MIYAKE, TINA M., Ph.D. Metacognition, Proactive Interference, and Working Memory: Can People Monitor for Proactive Interference at Encoding and Retrieval? (2007)
Directed by Dr. Michael J. Kane 118 pp.

The present study investigated whether subjects were sensitive to negative transfer and proactive interference (PI) at encoding and retrieval and whether sensitivity varied with working memory (WM) ability. Monitoring at encoding was assessed by having subjects make judgments of learning (JOLs; E1 & E2) or by controlling study time (E3) while learning word pairs. Monitoring at retrieval was assessed by dynamic prediction of knowing (DPOK) judgments. At encoding, the results suggest that subjects are sensitive to negative transfer at the list level but not the item level. At retrieval, subjects were sensitive to PI at the list level and sometimes at the item level. Sensitivity to negative transfer did not vary with WM, but sensitivity to PI did. Implications for control are discussed.

METACOGNITION, PROACTIVE INTERFERENCE, AND WORKING MEMORY:

CAN PEOPLE MONITOR FOR PROACTIVE INTERFERENCE

AT ENCODING AND RETRIEVAL?

by

Tina M. Miyake

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2007

Approved by

____
Committee Chair

APPROVAL PAGE

This dissertation has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair ___________________________________

Committee Members___________________________________

____

____

____
Date of Acceptance by Committee

__June 29, 2007______________
Date of Final Oral Examination

ACKNOWLEDGEMENTS

I would like to thank my advisor, Michael Kane, for his expertise, patience, and guidance on my dissertation. And, to the other members of my dissertation committee, John Dunlosky, Lili Sahakyan, and Stuart Marcovitch, thank you for your valuable expertise. I also would like to thank Douglas Levine, Julie Baker, Brad Poole, Jen McVay, and Michael Serra for their assistance in understanding gamma correlations and/or their input on the interpretation of the data. Finally, to Will Drath, Alisa Haymore, Shannon Holmes, Brad Kenyon, Kyle Lauricella, Alexis Lockett, Heather Motsinger, Jody Ortman, Amanda Page, Dara Rogers, Craig Rogers, Liz Simmons, Jacob Stewart, Lena Stuart, and Ashley Zimmerman, thank you for helping collect data

TABLE OF CONTENTS

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

Underwood (1945) argued that the single term, "proactive interference" (PI), was misleading because it referred to two separate phenomena: negative transfer at learning and proactive interference at recall. Negative transfer refers to when people learn information more slowly due to learning similar information previously, whereas PI refers to reduced recall of information due to similar items in long-term memory competing for access into short-term memory (Whitely, 1927; Underwood, 1945). This distinction is potentially important because variables that maximize performance during encoding (i.e., that reduce negative transfer) do not always maximize performance at a delayed test (i.e., that reduce PI; Schmidt & Bjork, 1992).

However, contrary to Underwood (1945), Schmidt and Bjork (1992) argued that learning and recall should not be separated because performance measurements at learning were ambiguous and that researchers should view long-term retention as *revealing* how much was really learned at encoding. In effect, they recommended that instructors keep in mind the long-term goal of retention by focusing on factors at encoding that would enhance long-term retention, even if it meant increased negative transfer. In order to illustrate their argument, Schmidt and Bjork (1992) cited a Landauer and Bjork (1978) experiment where subjects learned names and had to recall the last name when presented with the first. They found that increasing the interval between study and test (e.g. 0, 3, 9 intervening items) lead to slower acquisition (i.e. increased negative transfer) but more long-term retention (i.e. reduced PI) than did the condition where items were tested immediately after presentation.

Schmidt and Bjork (1992) argued that the additional difficulty in the expanding-interval condition prevented subjects from engaging in superficial massed rehearsal. The present research focused on why increased negative transfer at encoding would lead to better recall. The present research proposes that people might be able to enhance long-term retention by monitoring for negative transfer—becoming aware that learning has halted or slowed—in order to engage control processes to resolve negative transfer. Specifically, initial negative transfer could provide an opportunity for monitoring, leading to the use of better strategies as seen in the expanding-interval condition. The present research focused on whether people can actually monitor for negative transfer during learning and for PI during retrieval. Because the importance of monitoring relies on its coordination with control, the two are discussed together.

Monitoring and Control

Cognitive neuroscience researchers have recently argued that the role of monitoring in tasks needs to be better understood because monitoring processes might be necessary to recruit and modulate executive control processes (Botvinick, Braver, Barch, Carter, & Cohen, 2001). Furthermore, a better understanding of the relationship between monitoring and control of interference also has practical educational applications such as in the metacognitive literature.

In metacognition, the concept of monitoring and control being coordinated in order to accomplish a goal or task is not a new one. Nelson and Narens (1990), within a basic cognitive framework, proposed that people could monitor their progress in a task and use that monitoring to initiate and exert control. For example, a student studying for an exam might assess the degree to which he has learned the material in order to decide whether or not to keep studying the material. Nelson and Narens' proposal was not focused specifically on a particular problem like PI, but much of the research they reviewed concerned retrieving items from memory, which can be hindered by PI. In the case of PI, there is ample evidence demonstrating that subjects can control or reduce PI (e.g., Kane & Engle, 2000; Sahakyan & Delaney, 2003); however, there is a dearth of knowledge about how people *know* to exert that control.

Nelson and Narens (1990) argued that control and monitoring were coordinated in order to accomplish a goal. They proposed that goal-directed cognitive processes might be divided into at least two levels: an object-level and a meta-level. The object-level cognitive processes operate on objects or some other external stimuli. For example, when learning word pairs in a list, the processes responsible for associating the two unrelated words together, so that one word cues the retrieval of the other word, are object-level processes. In turn, meta-level cognitive processes operate on the object-level processes. So, if after list 1, a student thinks, "how well did I learn those word pairs?" she is making a meta-level judgment of learning (JOL). In making the JOL, the student is having a metacognitive experience because she is thinking about her own thinking.

At the meta-level, there is a goal state, which is a mental representation of the processes occurring at the object level (Nelson & Narens, 1990). In the above example, the student has a mental representation of the goal to memorize the material, as well as *how* to memorize it. If using a rote repetition strategy was not as beneficial as the student hoped, then she might use this monitoring information as justification for switching strategies. Thus, two information pathways called *control* and *monitoring* connect the object-level to the meta-level. *Control* refers to the flow of information from the meta-level to the object level, and *monitoring* refers to the flow of information from the object level to the meta-level (Nelson, 1996). In order for the goal to be accomplished, the current state at the object level must be transmitted back to the meta level, and, if the current state is judged as not progressing towards the goal state, adjustments (e.g., choose a new strategy, allocate more time, re-read previous material, focus attention) can be made to get back on track (Koriat & Goldsmith, 1996; Nelson & Narens, 1990). Thus, monitoring recruits the control process in the service of accomplishing a goal.

Monitoring, Control, and Working Memory

Working memory (WM) researchers have made similar arguments that monitoring and control are linked. Despite differing in their views of what specific abilities WM tasks reflect, Baddeley and Logie (1999) suggested that there was agreement amongst WM researchers (see Miyake & Shah, 1999) that WM reflects monitoring, processing, and maintaining information. WM is measured with different kinds of tasks, but most involve two components, storage and processing. The dual-component nature of WM tasks reflects the theoretical perspective advanced by Baddeley and Hitch (1974) that it is advantageous for people engaged in complex cognitive activities to be able to maintain information in active memory while processing relevant information (see Conway, Kane, Hambrick, Wilhelm, & Engle, 2005).

For example, a frequently used WM task is the operation span task (OSPAN; Turner & Engle, 1989). In OSPAN, the processing component is solving a math problem, and the storage component is remembering a word. The task consists of several sets of math problems and words varying in set size from 2 to 6. Subjects verify aloud whether the math problem is correct and then they say the word out loud. Subjects must therefore pay attention to the math problem in order to verify whether it is correct, but they cannot ignore the word because their eventual score depends on how many words they recall correctly. Subjects continue in this manner until they are cued to recall the words in the serial order that they saw them presented. For example, for a set size of 2, subjects would verify a math problem, say a word, verify a math problem, and say the word and then they would see the cue for recall. High scorers on OSPAN or other WM tasks are referred to as *high spans* while low scorers are referred to as *low spans*.

Recall differences between high spans and low spans typically emerge in high PI situations (Lustig, May, & Hasher, 2001; May, Hasher, & Kane, 1999), with high spans being better able to combat PI (Kane & Engle, 2000), perhaps by suppressing prior-list items (Rosen & Engle, 1998). Thus, WM seems related to control of interference. But is it also needed for monitoring? Converging research findings (Kane & Engle, 2000; Rosen & Engle, 1997; Rosen & Engle, 1998; Turley-Ames & Thompson, 2003) suggest that WM might be needed for monitoring in learning and memory. Thus, monitoring for PI might depend on WM resources, but before WM and monitoring are discussed, WM's role in PI control is discussed since WM's importance in combating PI is better understood.

WM and Control of Interference

Kane and Engle (2000) demonstrated that controlled processing was needed at both encoding and retrieval in order to combat PI. High and low span subjects (prescreened on OSPAN) learned 3 word lists that were drawn from the same semantic category, such as animals. After each list was learned, subjects performed a rehearsal-prevention task and then recalled the list. Span groups did not differ in mean recall for the first list, but on the 2nd and 3rd lists, low spans exhibited greater losses in comparison to their list 1 performance than did high spans, indicating that low spans experienced more PI buildup.

To assess whether high spans experienced less PI because they used controlled processing to combat it, Kane and Engle (2000) divided subjects' attention by having them repeatedly tap a complex finger sequence (index finger—ring finger—middle finger—pinkie) on the keyboard. Once the list-learning task began, subjects tapped the complex sequence during the encoding of each list, during retrieval of each list, or not at all. High span and low span subjects in the load-at-encoding and load-at-retrieval conditions exhibited equivalent proportional losses due to PI. That is, divided attention equated the two groups, specifically dropping high spans' performance to the level of low spans. When high spans were compared across the 3 experimental conditions, those in the two load conditions recalled fewer words on lists 2 and 3 (but not on list 1) than did those in the no-load condition. Thus, divided attention selectively increased high spans' vulnerability to PI.

In contrast, when low spans were compared across the 3 experimental conditions, there were no differences in PI. Kane and Engle (2000) argued that low spans used controlled processing to learn list 1, but exhausted their attentional capabilities in doing so. Thus, they could not use controlled processing to further combat PI present in the subsequent lists. Consistent with this, low spans under load recalled fewer words on list 1 than did low spans not under load, whereas high spans showed no load effect on list 1. In contrast, the increased PI exhibited by high spans under load suggests that they normally engage a control process during encoding and retrieval when faced with interference. Thus, Kane and Engle suggested that low spans might be more susceptible to PI because they cannot, or do not, use controlled processing to counteract it.

Rosen and Engle (1998) investigated whether high spans and low spans differed in PI control due to differences in the ability to suppress competing items. Subjects learned 3 lists consisting of semantically associated word pairs such as "Bird-Bath." In the interference condition, subjects saw the same cue (e.g. Bird) in the first and second list, but in the second list subjects saw the cue paired with a new, more weakly associated response (e.g. Bird-Dawn). The third list was the same as the first list (e.g. Bird-Bath). In the non-interference condition, subjects saw 3 different lists (e.g., Bird-Bath, Table-Salt, Candle-Stick). As predicted, high spans experienced fewer intrusions from list 1 during second-list learning compared to low spans, suggesting that high spans experienced less negative transfer at learning than did low spans from list 1. Moreover, high spans' escape from negative transfer was due to suppressing list 1 when learning list 2: high spans in the Interference-group retrieved list 1 responses more slowly than did those in the non-Interference-group when asked to retrieve list 1 responses (as list 3) after learning list 2. Furthermore, high spans were slower in retrieving list 1 responses (as list 3) than they were when retrieving list 1 responses the first time. In contrast, the low spans in the Interference-group were slightly faster than in the non-Interference-group, and they were not slower on retrieving list 1 responses (as list 3) when compared to their initial list 1 retrieval times. Thus, Rosen and Engle's data suggest that high spans suppressed list 1 responses when learning a related list 2.

WM and Monitoring

By suppressing list 1 responses, high spans might have been able to escape from negative transfer and PI whereas low spans could not. But how did high spans *know* to suppress first list items? Rosen and Engle (1997) proposed that high spans monitor for intrusions during retrieval tasks and then suppress them, whereas low spans exhaust their WM capabilities in monitoring alone, thereby leading to failed suppression and control. In the Rosen and Engle verbal fluency study, high spans and low spans generated animal names for 10 minutes out loud and were instructed to avoid repeating names. In addition to the fluency task, some subjects also had to track digits on a screen (a load-at-retrieval condition). In all experiments, high spans retrieved more unique animal names than did low spans. But, the addition of the digit-tracking task reduced generation of unique animal names only for the high spans (to the level of low spans), suggesting that only high spans used attentional-control processes to generate animal names. Rosen and Engle argued that low spans might not have used attention to retrieve animal names because a significant amount of their resources were directed towards monitoring for repetitions. And, indeed, when Rosen and Engle removed the monitoring component of the task by encouraging subjects to repeat animal names if they came to mind, low spans were more likely to resample animal names than were high spans, and low spans were just as likely to make a repetition as they were to retrieve a unique animal name.

However, Rosen and Engle's (1997) study only suggests that the ability to coordinate monitoring and control might depend on WM. They were not investigating the role of monitoring in retrieval. Thus, the link between control and monitoring has been implied but not demonstrated. That said, Rosen and Engle's argument is bolstered by a study by Turley-Ames and Thompson (2003) indicating that high spans might have better monitoring capabilities than low spans. Subjects read passages, and following each passage, predicted how well they would answer True/False questions about it. WM, as measured by OSPAN, correlated with performance on the True/False topic questions (r = .18), and with subjects' predictions about performance on topic questions (r = .53). Also, subjects' predictions about their future performance correlated with their actual performance (r = .29). More importantly, the relationship between WM and performance was mediated by subjects' predictions on how well they would answer the questions. That is, when the variation from subjects' predictions was partialed out, the correlation between WM and performance became nonsignificant. This implies that high spans' enhanced performance on the True/False questions depended largely on better monitoring. Turley-Ames and Thompson noted that the correlation between WM capacity and topic question performance was small, but their study, like Rosen and Engle's (1997), suggests that WM capacity might be needed for monitoring. Moreover, high spans are more likely to report using a strategy during OSPAN than were low spans, and they spent more time viewing the words in OSPAN (Turley-Ames & Whitfield, 2003) suggesting that high spans use strategies as a control process. Taken together, it is possible that high spans might use monitoring to recruit control (i.e., strategies) in situations like reading or in novel tasks like OSPAN, where PI can build up across trials. Although Turley-Ames and Thompson did not address PI, the use of better encoding strategies seems to help people escape from PI as demonstrated in directed forgetting (Sahakyan & Delaney, 2003).

Encoding Strategies and PI

Directed forgetting is seen when individuals are instructed to forget one set of materials, such as list 1, in favor of another set of materials, such as list 2 (Conway, Harries, Noyes, Racsma'ny, & Frankish, 2000; Sahakyan, 2004; Sahakyan & Delaney, 2003). One result of the instruction to forget list 1 is an increase in the recall of list 2 compared to subjects who were not instructed to forget (Bjork, 1970; Bjork, LaBerge, & LeGrand, 1968, Muther, 1965). This benefit of directed forgetting has been attributed to an escape from PI because subjects who have been instructed to forget list 1 (i.e. Forget group) show identical list 2 recall as subjects who learn only list 2. Sahakyan and Delaney (2003) demonstrated that this escape from PI could be attributed to better encoding strategies used on list 2 compared to list 1. Their subjects learned two lists of 15 words. On the first list, subjects were required to use a rehearsal strategy (i.e., shallow encoding), and on the second list, subjects made up a story with the words (i.e., deep encoding). If the benefit of directed forgetting is due to a switch to a better encoding strategy on list 2, then controlling for strategy should eliminate the benefit of directed forgetting, which is what Sahakyan and Delaney found.

Sahakyan, and Delaney (2003) investigated why subjects might switch strategies between list 1 and list 2. Specifically, they re-analyzed data from Sahakyan and Kelley (2002) and found that subjects in the forget group reported spontaneously switching strategies more often than did subjects in the remember group (i.e., subjects instructed to remember both lists). Sahakyan, Delaney, and Kelley (2004) proposed that the forget cue between list 1 and list 2 prompted subjects to evaluate the efficacy of their list 1 strategy (or lack thereof). In comparison, subjects who received the remember cue might not take the time to evaluate their strategy because it might cost them valuable time to rehearse the words. To test this idea, Sahakyan et al. had subjects learn two lists of 15 words. After list 1, half the subjects made a global judgment of learning (JOL), where they predicted the number of list 1 words they would be able to recall on the final test; the other half did not. Then, the forget or remember instruction for list 1 was given before subjects learned list 2. Subjects who were told to forget, and had a chance to evaluate list 1 by making a global JOL, did not differ in list 2 recall from subjects who were just told to forget list 1. In contrast, subjects who were told to remember list 1, and made a global JOL about list 1, recalled more list 2 words than did subjects who were just told to remember list 1. Sahakyan et al. argued that the normal tendency for subjects told to remember list 1 is to *not* evaluate the efficacy of their list 1 strategy.

It appears that subjects can evaluate their recall performance using a global JOL without actual feedback and switch to a better encoding strategy, thereby allowing them to escape from PI. What is still puzzling is what the evaluation is based on. One possibility is that the global JOL and other metacognitive judgments reflect subjects' awareness of interference in general.

Metacognition and Interference

Relatively few studies have investigated whether subjects use the presence of interference as a basis for JOLs. But, a few metacognitive researchers have demonstrated its influence in some contexts (Maki, 1999; Schreiber, 1998; Schreiber & Nelson, 1998). For example, Maki (1999) found that subjects were accurate in predicting their own performance in a traditional paired-associate task. Subjects learned two lists of number-word pairs (e.g., *321-rancher*) where the number served as the cue and the word served as the response. The goal on the subsequent cued recall task was to recall the word when shown the number. In the interference condition, the number cues from the first list were repeated in the second list and were associated with new words, and in the control condition, the number cues in the second list were completely new. After list 2 was learned, subjects made JOLs for the first list only and then took the cued-recall test for list 1. Maki found that control subjects gave higher JOLs to the number-word pairs than subjects in the interference condition.

However, subjects do not always use interference as the basis for their metacognitive judgments in paired associates tasks. Metcalfe, Joaquim, and Schwartz (1993) presented subjects with a list of word pairs such as "turtle-lucky." All subjects saw the same cue-target word pairs in the second half of the list. In the first half of the list, a cue was paired with a synonym of the response that would be seen in the second half of the list (e.g. AB'/AB: turtle-fortunate/turtle-lucky), or with the same response (e.g. AB/AB: turtle-lucky/turtle-lucky), or a new response (e.g. AD/AB: turtle-funny/turtle-lucky), or the cue was not repeated (e.g. CD/AB: lamp-funny/turtle-lucky). After studying the word pairs, subjects completed a cued-recall task for all the word pairs, and they were instructed to recall the second word associated with the cue if the cue was associated with more than one word. For the incorrect items (both omissions and commissions), subjects made feeling-of-knowing judgments on a scale from 1-100 about how sure they were that they would recognize the forgotten item on a subsequent recognition task. Then, subjects completed an 8-alternative forced-choice recognition task. In the first experiment, recognition was best in the AB/AB condition followed by AB'/AB, which did not differ from AD/AB, and lastly, CD/AB. Thus, subjects did not experience PI (AD/AB pairs were recalled better than CD/AB control pairs). In the second experiment, AB/AB was the best followed by CD/AB and then AB'/AB and AD/AB, which did not differ from each other, and so subjects did experience PI here. In both experiments, however, the feeling-of-knowing judgments were consistently based on cue repetition. When the cue was repeated, as in 3 of the conditions, subjects gave those word pairs a higher feeling-of-knowing compared to the condition without the repeated cues. This suggests that subjects based their feeling-of-knowing judgments on how familiar the cue was and not on what factors might influence recognition or cued recall performance, such as interference.

Another basis for people's judgments in interference situations is their beliefs about memory. For instance, McGuire and Maki (2001) had subjects learn the locations of different objects ("The exit sign is in the airport." "The exit sign is in the lounge."). Some locations contained more than one object (single-location, SL condition). For example, the *exit sign*, the *ceiling fan*, and the *coffee table* could all be located in the hotel lobby. As well, some objects could be located in more than one location (multiple-location, ML condition). After viewing the sentences, subjects completed a cued-recall test where they answered questions like, "What is located in the hotel lobby?" After studying the sentences in this way, subjects made JOLs for each on a 1-7 scale. Lastly, at test, subjects verified whether the presented sentence was studied or not (non-studied sentences were new combinations of the objects and locations in the studied sentences). Based on prior fan-effect findings (Radvansky & Zacks, 1991), it should take longer to recognize whether a sentence was studied or not in the ML condition because multiple representations need to be activated in order to verify whether a particular sentence was studied or not. In comparison, subjects should be faster in recognizing studied sentences in the SL condition because only 1 representation needs to be activated (Radvansky & Zacks, 1991). Indeed, sentence verification times were longer as objects were located in more places (i.e., in the ML condition), but not if they were located in one place. However, subjects gave lower JOLs as fan increased for *both* the SL and ML condition. The disassociation between verification times and JOLs in the SL condition suggests that subjects were not directly monitoring response competition. Thus, McGuire and Maki argued that subjects used a heuristic (e.g., a belief that 3 sentences will be harder to remember or recognize than 1 sentence regardless of integration) to make their judgments of learning.

In summary, subjects' metacognitive judgments can be based on more than one type of information (Koriat, Bjork, Sheffer, & Bar, 2004): 1) factors influenced by interference (Maki, 1999; Schreiber, 1998; Schreiber & Nelson, 1998); 2) beliefs about memory (McGuire & Maki, 2001); 3) cue familiarity, or amount of information retrieved (Eakin, 2005; Metcalfe et al., 1993). Consequently, metacognitive judgments in interference situations could be based on both the experience of interference and people's beliefs or knowledge about their own memory. The challenge will be to decipher what subjects are thinking when making these judgments.

All the interference and metacognition studies discussed involved metacognitive judgments made *after* an explicit recall attempt, even when they were made during the learning phase of the experiment. Consequently, these studies inform researchers about subjects' monitoring *after* retrieval but not about their monitoring *before* retrieval. As previously stated, the primary focus of the present research is whether people are able to coordinate monitoring and control to resolve PI. If people have a sense of difficulty while learning the material, then control can be exerted before a test rather than after a test. However, before research can be conducted on the possible link between monitoring and control in resolving PI, the question of whether people are influenced by negative transfer and PI at encoding and retrieval must be addressed in the first place.

When Experiment 1 (E1) was conducted, no other researchers to my knowledge had investigated whether metacognitive judgments about interference would be disassociated from recall performance if the judgments were made at encoding. However, since E1 was conducted, Diaz and Benjamin's (2005) unpublished research has come to our attention. Diaz and Benjamin also had subjects make judgments of learning at encoding before explicit retrieval attempts, and their research is discussed later in the Discussion. However, to preview what was found, Diaz and Benjamin's results confirm ours, that subjects might be aware of interference at encoding.

CHAPTER II

EXPERIMENT 1

E1 addressed whether people are sensitive to negative transfer and PI at encoding and retrieval, respectively. Subjects learned paired-associate lists while making immediate and delayed specific-item JOLs as an assessment of sensitivity to proactive interference. JOLs asked subjects to predict how likely it would be that they would recall the second word when presented with the first word on a scale from 0-99%. Immediate JOLs were made after each word pair, whereas delayed JOLs were made after the entire list was presented. Of particular relevance to the primary question of whether people are aware of interference at encoding are immediate JOLs. Immediate JOLs are more likely to be based on ease of learning an item (i.e., encoding fluency) rather than the ease of retrieving an item (i.e., retrieval fluency; Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Koriat & Ma'ayan, 2005), on which delayed JOLs can be based. Thus, if immediate JOLs were lower for the interference items on list 2 than the non-interference items on list 2, this would suggest that subjects are sensitive to PI.

For the metacognitive judgments at retrieval, we had subjects make dynamic prediction of knowing (DPOK) judgments (Vernon & Usher, 2003), with multiple, consecutive POKs for each item at retrieval. According to Metcalfe et al. (1993), subjects' metacognitive judgments at retrieval are based on cue familiarity. However, depending on how quickly the judgments must be made, the bases of metacognitive judgments may vary. If judgments must be made very quickly, they tend to be based on cue familiarity, but as more time is allowed to make metacognitive judgments, they tend to be based more on retrieveability (Benjamin, 2005; Koriat & Levy-Sadot, 2001). Consequently, during the retrieval process, consecutive POKs might vary as subjects either expect or experience greater difficulty in retrieving an item.

Finally, in order to induce proactive interference, half of the subjects learned two lists of Swahili-English vocabulary words (Interference-group: AD/AB) while the other half learned only one list (control group: AB only). For the interference subjects, there were two kinds of 2nd list items: interference (i.e. AD/AB) and non-interference (CD/AB) word pairs. The interference word pairs shared a cue across the two lists (e.g. tabibu—banker/tabibu—doctor), and the non-interference word pairs did not share a cue (nyanya—banker/ladha—flavor). Thus, interference and control subjects' list 2 recall were compared to assess whether there was a list-level PI effect. To assess whether PI was occurring at the item level for interference subjects, the recall accuracy for interference word pairs and non-interference word pairs were compared. If subjects can monitor for PI at the list level, then interference subjects should give lower JOLs than control subjects, and interference subjects should give lower JOLs for list 2 items than list 1 items. If subjects can monitor PI at the item level, then interference subjects should give interference items lower list 2 JOLs than non-interference items.

Method

Subjects

One hundred thirty-three native English speakers completed the study to fulfill part of their research requirement for their introductory psychology class. Data from 12 subjects were dropped: 9 subjects did not complete the study, 1 subject was ill, 1 spoke a language similar to Swahili, and 1 had learned Swahili before. Consequently, data from 121 subjects were included.

Design

There were 2 between-subject variables: JOL type and Interference-group. Subjects either made *immediate* item-specific JOLs, which occurred right after studying a word pair, *delayed* item-specific JOLs, which occurred after the entire list had been studied, or *no* item-specific JOLs. As described above, PI was induced at the list-level and item-level. Word pair type was a within-subject variable, and Interference-group was a between-subjects variable.

Materials

Thirty Swahili words were selected with their English translations from Nelson and Dunlosky's (1994) Swahili norms plus 10 six-letter English words, without their Swahili word counterpart, to construct the Swahili-English word pairs that would repeat (i.e. interference items). Swahili words with English translations consisted of 6 letters, and the normative mean likelihood of recall for the English word was matched across 3 recall trials (see Appendix A & B). Mean likelihood for recall is not available for the additional 10 English words because these words were de-coupled from their Swahili counterpart. The length of the Swahili word was not restricted, which ranged in length from 3 to 8 with the mode being 6 letters, and mean word length was 5.63.

Six lists were constructed with each consisting of 20 Swahili-English word pairs: 4 versions of list 1 (see Appendix A) and 2 versions of list 2 (see Appendix B). Because the structure of list 1 depended on the structure of list 2, I describe list 2 first. In list 2, the English word was the actual translation of the Swahili word. In order to avoid the possibility that certain Swahili-English word pairs might be easier than others to recall or learn, I counterbalanced Word pair type (i.e. interference vs. non-interference) across two versions of list 2. If a word pair was designated as an interference item on one version, then it was designated as a non-interference item on the other. There were four versions of list 1 because of the following (see Appendix A & B): First, because Word pair type was counterbalanced in list 2, at least two different versions of list 1 needed to be constructed to correspond to the two versions of list 2. However, since only half of the cues repeated in list 1, another set of Swahili-English word pairs needed to be selected to complete list 1. Thus, 4 versions of list 1 were constructed from three sets of Swahili cues and two sets of English translations. The 3rd set of Swahili cues were only learned in list 1, but which of the first 2 sets of Swahili cues used for the repeating Swahili cues in list 2 depended on which list 2 subjects learned: list 2 version 1 or list 2 version 2. Finally, the two sets of English translations were counterbalanced across the 4 versions of list 1.

Procedure

Subjects completed the study over two days. On the first day, subjects learned the lists (or if they were in the control condition, they completed a reasoning task and then learned one list, corresponding to list 2). Forty-eight hours later, subjects returned to complete the delayed recall test of list 2.

Session 1. Subjects in the Interference-group learned List 1 and completed an immediate recall test, and subjects in the control condition completed the Letter Sets task, which was adapted from the Educational Testing Service Kit of Factor-referenced tests (ETS: Ekstrom, French, Harman, & Derman, 1976). Then, all subjects learned the critical list 2 (the only list for control subjects) and completed an immediate recall test for list 2. Lastly, subjects were reminded to return 48 hours later for session 2.

List 1 learning proceeded in the following manner. Each Swahili-English word pair appeared on the screen for 4 s. Then, depending on the JOL condition, subjects either made an item-specific JOL (immediate condition), which asked them to judge how well they learned each word pair, or subjects studied the next word pair. After all the word pairs were presented, subjects in the immediate condition made a global JOL (The global JOL results are reported in the appendix in order to cut down on redundancy in the results section; see Appendix C). If subjects were in the delayed JOL condition, they were re-presented with all the Swahili cues in a random order and made item-specific JOLs. Subjects in the No JOL condition did not make specific-item JOLs. In making the JOLs, subjects were presented with only the Swahili word with the following question for the specific-item JOLs: *If 1 hour from now you were presented with the Swahili word, how likely would you be to remember the English word? Please make your rating on a scale from 00% (definitely will not remember) to 99% (definitely will remember)*. Subjects studied the list 3 more times with word pairs presented in a random order. Then, in the self-paced immediate recall test for List 1, subjects saw each Swahili word and typed in its English translation. Subjects were allowed to guess or leave it blank. The order of the Swahili words was random, and subjects were not penalized for guessing or misspellings. List 2 was learned in the same manner as List 1, except List 2 was presented 3 times for the interference subjects and 4 times for the control subjects. Because interference subjects might learn list 2 better because of the benefit of learning-to-learn (Postman, 1972), list 1 was presented 4 times to the control subjects in order to equate learning on list 2. An immediate recall test followed learning of list 2. Again, subjects were not penalized for guessing or misspellings.

Subjects in the control condition learned only one list, which corresponded to List 2 in the Interference-group. Rather than learn list 1, then, the first task that controls completed was the Letter Sets task. The Letter Sets task lasted approximately 13 minutes, which was the approximate time that most subjects took to complete list 1 learning in pilot studies. After completing the Letter Sets task, subjects learned the list of Swahili-English word pairs in the same manner as subjects in the Interference-group except controls saw the list 4 times.

The Letter Sets task consisted of 20 problems. On each trial, subjects saw five letter sets with 4 letters each. Four of the five letter sets were constructed based on a common rule. For instance, four of the sets might contain the letter “L” where the remaining letter set did not. Subjects had to select the letter set that was different and press the corresponding key (1, 2, 3, 4, or 5). Each problem appeared on the screen for 33 seconds, and subjects had 5 seconds to enter their response. Afterwards, subjects made a confidence judgment: "How confident are you that you selected the correct answer? Please make your rating on a scale from 00% to 99%."

Session 2. Forty-eight hours later, subjects returned to complete session 2. The first task that all subjects completed was another version of the Letter Sets task, but this time, subjects heard a 150 ms tone prompting them to make a judgment. The purpose of this version of the Letter Sets task was to familiarize subjects with the DPOK procedure that would be used during the subsequent delayed recall test of list 2. Subjects heard the tone four times, and the tones were 5 s apart. Every time subjects heard the tone they made a judgment as to how close they were to solving the problem: 20%, 40%, 60%, or 80% close to solution. Experimenters instructed subjects to make their judgments as quickly as possible since their judgments were supposed to reflect their “gut feeling.” A strict RT deadline was not imposed, but in order for the computer to record the response, subjects had to respond within the 5-s interval. Thus, experimenters told subjects not to pause for more than a second or so to make their judgments. After making 4 judgments, subjects entered their solution to the problem. Subjects had 4 s to enter their solution to the problem before making the confidence judgment. The Letter Sets task consisted of 10 problems. None of the 10 problems overlapped with the problems from controls’ session 1.

After subjects finished the Letter Sets task with the tones, subjects completed the delayed recall test via the same DPOK procedure. Each Swahili word from list 2 appeared on the screen, without its English translation, in random order. Subjects heard a tone every 5 s and judged how close they were to recalling the English translation (from list 2 for interference subjects), and 5 s after the 4th DPOK prompt, subjects typed in their answer. If they could not recall the English word, subjects were allowed to guess or press the "ENTER” key to move on to the next word. Subjects had as much time as they needed to recall the word.

Results

The main findings concern the metacognitive judgments at encoding and retrieval. The primary question I am interested in is whether people are sensitive to PI at encoding, especially at the item-level. First, however, I present the recall data to assess PI. For the inferential statistics reported here, I set the alpha at .05 and effect sizes are reported as partial eta squared (η_p^2).

List 2 Recall Accuracy

A 2 (Recall Test: immediate, delayed) x 2 (Interference-group: control, interference) x 2 (Word pair type: non-interference, interference) x 3 (JOL type: immediate, delayed, no JOL) repeated measures ANOVA was conducted on the recall performance for list 2. Interference-group and JOL type were between-subjects variables, and Recall test and Word pair type were within-subjects variables.

Subjects recalled more words at the Day 1 immediate test ($M = .61$, $SD = .25$) than the Day 2 delayed test ($M = .43$, $SD = .25$), $F(1, 115) = 233.60$, $\eta_p^2 = .67$, and control subjects ($M = .58$, $SD = .24$) recalled more words than interference subjects (M = .46, SD = .24), $F(1, 115) = 8.41$, $\eta_p^2 = .07$. However, both main effects were qualified by a significant interaction between Interference-group and Recall test, $F(1, 115) = 15.15$, $\eta_p^2 = .12$, indicating a greater PI effect at the delayed test than at the immediate test (see Figure 1). Two independent t-tests were conducted to follow-up the interaction. At immediate recall, control subjects recalled more English words than did interference subjects, but the difference was only marginally significant, $t(119) = 1.82$, $p = .07$. At delayed recall, control subjects recalled significantly more words than interference subjects, t (119) = 3.84. Consistent with Postman, Stark, and Fraser (1968), then, the list-level PI effect increased over time and was only significant after the delay.

Although interference subjects differed from control subjects overall, indicating a list-level PI effect, we did not find an item-level PI effect for interference subjects. That is, there was no interaction between Interference-group and Word pair type, $F < 1$, $p = .45$, nor was there an interaction between Recall test, Interference-group, and Word pair type, $F < 1$, $p = .49$.

JOL type was included in the ANOVA because it was one of the manipulated variables; however, JOL type was not expected to be a significant factor in influencing recall performance. Instead, JOL type was manipulated because Dunlosky and Nelson (1992) found that JOLs made after a delay rather than immediately after the item were more accurate in predicting future recall. As predicted, subjects in all 3 JOL type groups performed equally well on the recall tests, $F < 1$, $p = .75$. (Remaining significant effects were two-way interactions between JOL type and Recall test, F (2, 115) = 6.91, $\eta_p^2 = .11$, and Word pair type and Recall test, F (1, 115) = 6.33, $\eta_p^2 = .05$. No other effects were significant.)

In summary, interference subjects recalled significantly fewer words than did control subjects at delayed recall, indicating a list-level PI effect. However, interference subjects did not recall fewer interference items than non-interference items, indicating no item-level PI effect.

Metacognitive Judgments at Encoding

The two primary questions about metacognitive judgments and negative transfer at encoding were the following: First, would subjects who learned two lists (i.e., Interference-group) give lower specific-item JOLs to list 2 than would subjects who learned one list (i.e., control group)? Second, would interference subjects experience the interference and non-interference items differently? Despite not obtaining the item-level PI effect, I kept Word pair type in my analyses because subjects might still have experienced the interference items differently from the non-interference items. Recall that Metcalfe et al. (1993) found that metamemory judgments did not necessarily follow the same pattern as recall performance.

List 2 specific-item JOLs. Subjects made specific-item JOLs for each word pair. A 2 (Interference-group: control, interference) x 2 (JOL type: immediate, delayed) x 2 (Word pair type: non-interference, interference) x 3 (Trial: 1^{st}, 2^{nd} 3^{rd} presentation of list) repeated-measures ANOVA was conducted on the mean JOL magnitudes for list 2. Interference-group and JOL type were between-subjects variables, and Word pair type and Trial were within-subjects variables. Figure 2 presents mean JOL magnitude for the control and interference subjects across the 3 learning trials.

Subjects gave significantly higher JOLs at each presentation of list 2, F (2, 152) = 77.34, $\eta_p^2 = .50$. Control subjects gave higher JOLs than did interference subjects, F (1, 76) = 6.05, $\eta_p^2 = .07$, suggesting that interference subjects were sensitive to the overall difficulty they were experiencing at encoding due to having learned list 1. The lack of an item-level effect (i.e., the comparison of JOLs for interference and noninterference items) was consistent with the recall data, where interference subjects recalled the interference and non-interference items equally well.

(Other significant effects were the following: Trial and JOL type, $F(2, 152) = 8.40$, $\eta_p^2 = .10$ and Word pair type x Trial x JOL type, $F(2, 152) = 4.12$, $\eta_p^2 = .05$. No other effects were significant.)

If interference subjects were sensitive to negative transfer, then their list 2 JOLs should have been lower than their list 1 JOLs. A 2 (JOL type: immediate, delayed) x 2 (List: 1, 2) x 2 (Word pair type: non-interference, interference) x 3 (Trial: 1, 2, 3) repeated-measures ANOVA compared interference subjects' list 1 JOLs and list 2 JOLs. JOL type was a between-subjects variable, and List, Word pair type, and Trial were within-subjects variables. Consistent with the PI sensitivity hypothesis, interference subjects gave higher JOLs for the first list than the second list, $F\ (1, 40) = 8.07$, $\eta_p^2 = .17$ (see Figure 3). JOLs also increased across trials, $F\ (2, 80) = 48.44$, $\eta_p^2 = .55$. Three paired t-tests with a Bonferroni correction indicated that subjects gave significantly higher JOLs at each subsequent learning trial. The remaining significant effect was an interaction between JOL type and trial, $F(2, 80) = 4.38$, $\eta_p^2 = .10$. (No other effects were significant.)

List 2 JOL magnitude also was analyzed by dividing the interference items into two groups: those recalled correctly during first list immediate recall and those not recalled correctly on first list immediate recall. Subjects might have noticed that the interference word pairs had two responses only if they recalled the first response on the immediate recall test; thus, subjects might have given lower list 2 JOLs only if they recalled the first list response correctly on the first test (M. Serra, personal communication, May 4, 2006). If so, then this would be evidence that subjects used experience gained from the first list (i.e. learning list 1 and/or taking the immediate recall test of list 1, which did not provide explicit feedback about correctness) to inform their judgment of list 2 learning. Consequently, I coded the interference word pairs on list 2 into 3 groups: interference/recalled on list 1, interference/not-recalled on list 1, and non-interference.

A 2 (JOL type: immediate, delayed) x 3 (Trial: 1, 2, 3) x 3 (Item: not-recalled interference, recalled interference, non-interference) repeated-measures ANOVA was then run on list 2 JOL magnitude. There was a main effect of item, $F(2, 78) = 13.26$, $\eta_p^2 = .25$. Subjects gave higher list 2 JOLs for interference word pairs that they had correctly recalled the first list response on the immediate recall test ($M = 37.88$, $SD = 20.18$) than interference word pairs where the first list response was not correctly recalled ($M = 27.20$, $SD = 15.09$), $t\,(40) = 4.37$, and non-interference word pairs ($M = 30.32$, $SD = 14.96$), $t\,(41) = -3.42$. The only other significant effect was an interaction between JOL type, Item, and Trial, $F(4, 156) = 3.12$, $\eta_p^2 = .07$. Because the significant main effect of item might have been due to differences in total list 1 recall, I re-ran the analyses with list 1 accuracy as a covariate. The main effect of item was still significant, $F\,(2, 76) = 13.65$, $\eta_p^2 = .26$, indicating that the effect of item was not spurious.

In conclusion, the JOL data suggest that interference subjects are sensitive to the overall negative transfer present at encoding. But, interference subjects were not differentially sensitive to negative transfer between interference and non-interference items. Furthermore, interference subjects might have used the experience of list 1-learning and recall when making their list 2 JOLs.

Gamma correlations. I calculated gamma correlations between list 2 JOLs and list 2 immediate and delayed recall performance. Gamma correlations are used as a measure of whether or not subjects' metacognitive judgments can discriminate between items that will or will not be recalled later. High, positive gamma correlations indicate that subjects gave higher judgments to recalled items than they did to unrecalled items, whereas high, negative gamma correlations indicate the reverse. Gamma correlation around zero indicate that subjects were assigning higher judgments to recalled and unrecalled items equally. Unlike the recall performance data, I did expect JOL type to influence gamma correlations since previous researchers have found that delayed JOLs are more accurate than immediate JOLs (Dunlosky & Nelson, 1992).

For each presentation of the list, a gamma correlation was calculated between JOLs and recall. For instance, a subject's JOLs, made during each presentation of list 2, were each correlated with that subject's immediate recall test performance of list 2. After obtaining the individual subjects' gamma correlations, I conducted ANOVAs on the mean gamma correlations to ascertain whether there were group differences in relative accuracy.

A 2 (Interference-group: control, interference) x 2 (JOL type: immediate, delay) x 2 (Word pair type: non-interference, interference) x 3 (Trial: 1, 2, 3) x 2 (Recall test: immediate, delayed) repeated-measures ANOVA was calculated on the individual gamma correlations between list 2 JOLs and recall of list 2. I found no significant effects.

Most means were relatively low or negative indicating that subjects' list 2 JOLs did not discriminate between recalled and unrecalled items (see Figures 4 & 5).

Metacognitive Judgments at Retrieval

Subjects completed the delayed recall task via the DPOK procedure, which required making 4 consecutive judgments prior to giving a recall response. Subjects made each judgment quickly, and 5 s passed between each. I predicted that cue familiarity might influence the first DPOK judgment, and that retrieval fluency, influenced by PI, might influence subsequent judgments.

DPOK judgments. A 2 (Interference-group: control, interference) x 3 (JOL type: immediate, delayed, No JOL) x 2 (Word pair type: non-interference, interference) x 4 (DPOK prompt: 1, 2, 3, 4) repeated-measures ANOVA was conducted on the mean magnitude of DPOK judgments (see Figure 6). Interference-group and JOL type were between-subjects variables, and Word pair type and DPOK prompt were within-subjects variables. DPOKs increased across trials, F (3, 345) = 142.88, η_p^2 = .55. Six paired t-tests with a Bonferroni correction were conducted as follow-up tests. Subjects gave significantly higher DPOK judgments at each successive DPOK prompt.

Subjects in the control condition gave higher DPOK judgments than did subjects in the Interference-group, suggesting that interference subjects were sensitive to PI present during the delayed recall test, F (1, 115) = 6.79, η_p^2 = .06. Furthermore, a marginal interaction between Interference-group and Word pair type was found, F (1, 115) = 3.28, p = .07, η_p^2 = .03. Subjects in the Interference-group gave slightly higher DPOK judgments to the interference items than to the non-interference items, t (61) = -

1.64, $p = .11$, and subjects in the control condition gave equivalent DPOK judgments to the non-interference and the interference items, $t < 1$, (see Figure 6). This finding, albeit weak, suggests that interference subjects were influenced by cue familiarity at retrieval. (The only other significant effect was an interaction among JOL type, Interference-group, and DPOK prompt, $F(6, 345) = 2.43$, $\eta_p^2 = .04$.)

Interference subjects' difficulty with cue familiarity could have been due to the requirement to make rapid judgments. However, an interaction between Interference-group and DPOK prompt was not found, $F < 1$, $p = .78$, suggesting that even after two or three prompts (approx. 10-15 s), interference subjects were still being influenced by cue familiarity and giving higher judgments to interference items than non-interference items. Word pair type did not interact with any of the other variables, $F < 1$, $p > .70$, which was consistent with the lack of an item-level PI effect in the actual recall performance. Because interference and non-interference items were recalled equally well by interference subjects, the items should not have been rated differently.

Gamma correlations between DPOK judgments and delayed recall. Gamma correlations were calculated between DPOK judgments and delayed recall of list 2. A 2 (Interference-group: control, interference) x 3 (JOL type: immediate, delay, no JOL) x 2 (Word pair type: non-interference, interference) x 4 (DPOK prompt: 1, 2, 3, 4) repeated-measures ANOVA was calculated on the individual gamma correlations (see Figure 7). A significant main effect for DPOK prompt was found, $F(3, 252) = 7.69$, $\eta_p^2 = .08$, but, after conducting paired t-tests with a Bonferroni correction, significant pairwise comparisons were not found.

DPOK prompt interacted with Interference-group, F (3, 252) = 2.82, η_p^2 = .03, and Word pair type interacted with Interference-group, F (1, 84) = 5.76, η_p^2 = .06. However, a 3-way interaction among DPOK prompt, Interference-group, and Word pair type qualified the 2-way interactions, F (3, 252) = 2.68, η_p^2 = .03 (see Figure 7). I therefore conducted two separate 2 (Word pair type: non-interference, interference) x 4 (DPOK prompt: 1, 2, 3, 4) repeated measures ANOVAs for control and interference subjects. Control subjects showed no significant effects, being equally accurate in discriminating items they were about to recall or not across the DPOK prompts, regardless of whether the item was an interference or non-interference item. However, interference subjects more accurately discriminated whether a non-interference item would be recalled than they did for interference items, F (1, 44) = 4.06, p = .05, η_p^2 = .08. A main effect for DPOK prompt was found, F (3, 132) = 6.28, η_p^2 = .12: Interference subjects were more accurate at the 4th DPOK prompt than at the 1st DPOK prompt, t (58) = -2.91. This effect could have been due to the interference cues being more familiar, and if by the 4th prompt the item were not retrieved, this would have provided explicit proof that they were not likely to recall the item. No other comparison was significant at the .008 level (Bonferroni correction).

Subjects were very accurate in predicting whether they were about to recall the word or not as evidenced by the high gamma correlations between DPOK judgments and delayed recall of list 2. Control subjects gave significantly higher DPOK ratings than interference subjects, and their judgments were consistently more accurate than interference subjects. Moreover, interference subjects' judgments were not as accurate for interference items as they were for non-interference items.

Discussion

E1 assessed whether people were sensitive to negative transfer and PI, at the list and item levels, for encoding and retrieval. List-level PI effects were significant: Control subjects recalled more words than interference subjects at the delayed recall test. Consistent with the idea that people's judgments are sensitive to negative transfer at encoding, control subjects also gave higher JOLs to the word pairs than interference subjects. Moreover, interference subjects gave lower JOLs for list 2 than list 1, suggesting that they were sensitive to the list-level negative transfer present at encoding. Unfortunately, item-level PI effects were not significant: Interference subjects recalled the interference and non-interference items equally well, and consistent with this, interference subjects did not give higher JOLs to the non-interference items than the interference items.

On one hand, the negative transfer/PI sensitivity explanation suggests that subjects based their judgments on the experience of negative transfer and PI. On the other hand, without the item-level PI effect and the predicted difference between interference and non-interference items, the pattern of results can be explained from a theory-based perspective, which is consistent with Diaz and Benjamin's (2005) unpublished findings.

Diaz and Benjamin (2005) used a similar PI buildup and release design that Kane and Engle (2000) did. Subjects learned 4 lists of word pairs. The first 3 lists were constructed so that the cues repeated across the lists (AB, AC, AD lists), and the 4th list was the PI release list so it consisted of entirely new cue-response pairs (EF list). Subjects made JOLs after studying each item. Diaz and Benjamin found that subjects' JOLs decreased across all 4 lists despite recall improving on the 4th list; thus, subjects were insensitive to PI release. In their second study, Diaz and Benjamin had subjects learn 7 lists where the 4th and 7th lists were the PI release lists. They hypothesized that subjects might have unsophisticated mental models about their memory that were not updated in time when they experienced the first PI release. Thus, if they had a second chance to make JOLs during a PI release list, then maybe their JOLs would increase rather than decrease. Again, subjects exhibited the decrease in recall across the PI buildup lists and an increase in recall on the PI release lists; however, subjects' JOLs on the 4th and 7th lists did not increase but on the 7th list JOL magnitude appeared to level out. Consequently, Diaz and Benjamin argued that the JOLs were based on theory rather than mnemonic cues like encoding fluency. Thus, in the present study, subjects in the Interference-group might have had a theory that the second list would be harder to learn due to interference and lowered their JOLs accordingly. Although this is possible, it seems equally likely that subjects might have believed that list 2 should be easier to learn because they had experienced a similar learning situation with list 1.

An even more likely possibility is that experience of learning list 1 was more influential than a prior theory about PI, because theory-based JOLs are difficult to elicit from people when they are making judgments about themselves (Koriat, et al., 2004). Consistent with this, interference subjects gave higher JOLs to list 2 items where the cue was previously paired with a list 1 item they had recalled previously than to non-interference items or interference items that were not recalled on list 1. It is possible that their success in recalling the correct response associated with recalled interference cues lead them to be overconfident. Thus, interference subjects might have been using their experience in learning and recalling list 1 to inform their JOLs. However, this poses a problem in interpreting the results. If interference subjects were influenced by their list 1 experience, how do we know that their list 2 JOLs reflect a sensitivity to negative transfer or lack of negative transfer at encoding? E2 addressed this problem.

At retrieval, subjects were sensitive to PI, consistent with previous research (Maki, 1999; Schreiber, 1998; Schreiber & Nelson, 1998): Control subjects gave higher DPOKs during delayed recall than did interference subjects. Furthermore, interference subjects tended to give higher judgments to interference items than to non-interference items at retrieval, suggesting that they were basing their judgments on familiarity with the interference cues. Moreover, interference subjects' judgments were less accurate on the first DPOK prompt than on the fourth DPOK prompt, suggesting that initially they were basing their judgments on cue familiarity, but that they were able to correct this by the 4th prompt. In contrast, control subjects' judgments were consistently accurate. The gamma correlation results are inconsistent with Maki (1999), who found that subjects' gamma correlations were not influenced by interference manipulations. The difference between Maki's study and the present research was probably due to procedural differences: Maki used a between-subjects manipulation to induce retroactive interference (RI), which is when recently learned items interfere with previously learned items. If a subject was in the interference condition, he/she learned all interference items for list 2. When they made judgments about list 1, cue familiarity should have been relatively equal across items. In the present research, in addition to the between-subjects PI manipulation, a within-subjects manipulation was used. Interference subjects might have relied on cue familiarity because that was the only diagnostic cue available, whereas control subjects had only their experience of recalling items on which to base their POKs. The DPOK data are consistent with Benjamin (2005) who found that judgments that are made quickly tend to rely more on cue familiarity than they do on retrieveability.

In summary, E1's results support the contention that people can be sensitive to negative transfer and PI at encoding and retrieval, respectively. However, there are still unanswered questions. By itself, the result that interference subjects gave lower JOLs for list 2 than control subjects is ambiguous because explanations other than the PI/negative transfer-sensitivity hypothesis are viable, as described above. The ambiguity of E1 results underscores the necessity of obtaining an item-level effect of PI in Experiment 2 (E2).

CHAPTER III

EXPERIMENT 2

E2 addressed three primary methodological difficulties with E1. First, the immediate recall test for list 1 was eliminated, because recalling list 1 might have influenced subjects' JOLs in the Interference-group when encoding list 2. If so, the decrease in the Interference-group subjects' JOLs from list 1 to list 2 might have reflected retrieval fluency during immediate recall of list 1 rather than the encoding fluency of list 2. Second, in order to maximize the chances of obtaining an item-level PI effect, the AB/AD design was replaced with the AB/ABr design because the AB/ABr design maximizes negative transfer (Postman, 1972). Thus, list 2 consisted of 8 new word pairs and 8 word pairs that were re-combinations of cues and responses from list 1. Finally, the Swahili-English word pairs were replaced with concrete, English noun-noun word pairs.The Swahili-English word pairs might have been easier to memorize than English, noun-noun word pairs; consequently, the difficulty in inducing PI in E1 might have been partially due to the Swahili-English word pairs. Also, JOL type was not manipulated in E2 since there was no effect in E1.

E2 also was conducted to investigate the role of working memory capacity (WM) in people's ability to monitor for PI. As mentioned in the introduction, despite not differing on list 1 recall in a PI buildup & release design, high spans remember more words than low spans on subsequent PI buildup lists (Kane & Engle, 2000).

Consequently, in E2, subjects were pre-screened on three WM span tests: operation span, reading span, and spatial span (Kane et. al, 2004).

I predicted that if WM was important for monitoring negative transfer in order to exert control (e.g., inhibition), then high spans should be better able to monitor for negative transfer than low spans: For high spans, list 2 JOLs should be lower in the interference condition than in the control condition. If low spans have poor monitoring ability, then low spans in the interference and control conditions should not differ in their specific-item JOLs. Moreover, high spans' specific-item JOLs should better discriminate between words that were recalled and words that were not recalled than low spans' JOLs.

Because people's JOLs might depend on either the experience of doing the task, or their beliefs about memory, or both (Koriat et al., 2004), I also assessed people's beliefs about interference in E2. Koriat et al. found that subjects accurately predicted that increasing the retention interval between study and test would decrease memory accuracy in *other* subjects, suggesting subjects based their predictions on a belief about how time and memory are related rather than basing their predictions on mnemonic or experiential cues since those cues were not available. In contrast, when people make predictions about themselves, Koriat et al. argued that they use a mixture of both theory and experience to make metacognitive judgments. Thus, subjects who completed the memory experiment did not seem to take into account the effect of retention interval on their own memory when making JOLs. Because no one I am aware of has investigated what people believe about memory and interference, I assessed whether people had any beliefs about interference at all by presenting them with a memory interference experiment and asked them to predict the recall rates for a group learning 1 list vs. a group learning 2 lists. The description of the experiment was E1 without the JOLs.

Method

Subjects

One hundred thirty-three subjects were recruited from the introductory psychology pool at UNCG. All subjects were required to have participated in a prior study where they completed 3 WM tasks and the memory beliefs questionnaire. One subject did not complete this prior study and was excluded from the analyses. Data from one subject were excluded because he completed the study with a sign-language interpreter, which could have influenced the results. Finally, data from one subject were excluded because she was not trying to do the study and disregarding experimenter instructions.

Design

Interference-group (interference vs. control) was the between-subjects variable. Word pair type (interference vs. non-interference items) was a within-subject variable. For the interference subjects, List (list 1 vs. list 2) was a within-subjects variable, and WM was a covariate.

Materials

Two versions of list 1 and list 2 were constructed from word pairs drawn from Hertzog , Kidder, Powell-Moman, and Dunlosky (2002) (see Appendix D). The word pairs in both versions of list 2 (critical list) were the same. The two versions of list 2 differed in which word pairs would be designated as interference and non-interference items. If in version 1, a word pair was an interference pair, then in version 2, it was not. This rule also was used for list 1. To construct list 1, the cues and responses from the interference word pairs in list 2 were re-combined. To determine which cues would be re-combined with which responses, each word pair in list 2 was assigned a number from 1 to 8. The cues remained in the same order from 1 to 8. A random number generator was used to re-order the responses. For version 1, the order of responses was 8, 5, 1, 6, 2, 3, 4, and 7, and in version 2, the order of the responses was 3, 5, 8, 6, 4, 1, 2, and 7. For example, for version 1, cue #1 (i.e. bone) was re-paired with the response from word pair #8 in list 2 (i.e. dancer), and cue #2 (i.e. animal) was re-paired with the response from word pair #5 in list 2 (i.e. senator).

The memory beliefs questionnaire was modeled after Koriat et al. (2004) and consisted of 2 memory scenarios: Retention intervals (Q1) and the AB/AD design (Q2) (see Appendix E). The retention intervals question was a filler question. The critical question was the AB/AD design question, which was analogous to the procedures from E1. Subjects predicted the percentage of interference items from list 2 that subjects in the hypothetical experiment would recall. They made a similar prediction for the non-interference items from list 2.

Working Memory Screening

In a prior session, subjects completed 3 automated WM tests in groups of 1 to 6: operation span (OSPAN), spatial span (SSPAN), and reading span (RSPAN), in that order. All 3 WM tests required subjects to make a processing judgment and then remember a to-be-remembered item. Subjects needed to maintain at least 85% accuracy on the processing judgments to be eligible to participate in the list-learning phase of the experiment.

In OSPAN, subjects saw a series of math problems paired with a letter to be remembered. A math problem (e.g., (2 * 2) + 1 = ?) appeared on screen followed by a possible solution on the next screen (e.g. 6). Subjects clicked "true" or "false" that the math problem solution was correct. If subjects did not solve the math problem and click to the solution screen quickly enough, the computer automatically skipped to the letter, and the problem was counted as incorrect. The time limit was calculated for each subject (mean + 2.5 SDs) during the math practice. After clicking true or false, a letter appeared on screen for 1 s. Then, another math problem appeared followed by another letter. At the end of each set, a recall screen appeared with 12 letters. Subjects had to click inside the box of each letter in the order they had appeared. If they could not remember a letter, they had to click the "blank" box as a place-holder for that letter.

In RSPAN, subjects saw sentences that either made sense (e.g. During the winter, you can get a room at the beach for a very low rate) or not (e.g., Andy was stopped by the policeman because he crossed the yellow heaven). Subjects clicked "true" or "false" that the sentence made sense. As in OSPAN, subjects had to respond to the sentence within a time limit determined by their reading-only practice. Subjects saw the same 12 letters from the OSPAN on the recall screen. All other procedures matched OSPAN.

In the SSPAN, subjects saw a matrix pattern of black and white squares that was vertically symmetrical or not. Subjects clicked "yes" or "no" to as to whether the figures were symmetrical within the time limit determined from the symmetry judgment-only practice. After making their judgment, a red square appeared within a 4 x 4 grid for 650 ms, and subjects had to remember the location of that red square in the grid. Then, another symmetry figure appeared. At the recall screen, subjects clicked inside a 4 x 4 square grid in the order in which the red boxes appeared.

Scoring. In the OSPAN and RSPAN, a trial consisted of one math problem or one sentence followed by a letter. There were 15 sets of trials with set size ranging from 3 to 7 letters. There were 3 sets of each of the 5 set sizes. A letter was scored as correct only if it was in the correct position. For example, if a subject recalled all 7 letters in the set but the order of the last two were incorrect, the subject would receive 5 out of 7 correct for that set. The highest score possible was a 75.

In the SSPAN, a trial consisted of one figure followed by one red square. There were 12 sets of trials with set size ranging from 2 to 5 red squares. There were 3 sets of each of the four set sizes. A red square was scored as correct only if it was in the correct location and order that it appeared. For example, if a subject recalled 3 red squares out of four but only 1 was in the correct order, then the subject would receive 1 out of four correct. The highest score possible was a 42.

After completing the WM tests, subjects filled out the memory beliefs questionnaire.

List-learning procedure

Subjects in the Interference-group learned a list of unrelated noun-noun word pairs presented for 4 s each. Immediately after list-1 learning, the Interference-group subjects learned a second list of unrelated noun-noun word pairs presented for 4 s each, followed by an immediate cued-recall test of list 2. Subjects in the control condition completed the same Letter Sets task from E1 and then learned the same list 2 as subjects in the interference condition.

While learning the list(s), all subjects made immediate JOLs. In making the JOLs, subjects saw only the first word from the word pair, and they answered the same JOL question from E1. After all the word pairs appeared on screen, subjects answered the same global JOL question from E1.

After learning the list(s), subjects completed two filler tasks for 15 minutes. A pilot study was conducted to determine the length of the delay interval. The first task had subjects make familiarity judgments about artwork presented on a computer screen. The second filler task was same Letter Sets task as in E1, with one exception. Instead of hearing 4 tones during each trial, subjects heard 2 tones 10 s apart. The first tone occurred 5 s into the trial as in E1, and the second tone occurred 15 s into the trial. After subjects finished the Letter Sets task, they completed the delayed recall test for list 2 via the DPOK procedure as described above. Subjects saw the first word of the unrelated word pair and each time they heard the tone they made a judgment as to how close they were to recalling the second word.

Results

WM

The mean scores for OSPAN, RSPAN, and SSPAN, respectively, were: 50.09 (*SD* = 15.76), 45.16 (*SD* = 15.96), and 25.67 (*SD* = 8.28). OSPAN significantly correlated with RSPAN at .54 and with SSPAN at .56. RSPAN and SSPAN significantly correlated with each other at .50. Because the variance in a WM test can be attributed to several sources (e.g., spatial ability, verbal ability, fatigue, motivation, etc.) besides the construct of interest (e.g. working memory ability), any one WM test is not a perfect representation of the latent construct. Thus, a z-score WM composite created from 3 different WM tests is a better representation of the latent construct than a single test. Consequently, the raw scores were transformed into z-scores and averaged to create a z-score WM composite for each subject. The WM composite was entered as a covariate in subsequent analyses (see Oberauer, 2005).

Memory Beliefs Questionnaire

Here, subjects predicted the recall performance of hypothetical subjects in an interference condition, having been given the percentage correct of the control group. A 2 (Interference-group: interference, control) x 2 (Word pair type: interference, non-interference) repeated measures ANCOVA with WM as the covariate was conducted. Subjects predicted that the hypothetical subjects in the experiment would recall more interference items (M = 37.29, SD = 16.36) than non-interference items (M = 33.80, SD = 18.07), F (1, 122) = 4.74, η_p^2 = .04. This is not consistent with the proposal that subjects might believe that an item associated with two responses would be more difficult to recall than an item with one response. It also is not consistent with the actual recall results of E1 where there was no difference between interference and non-interference items. No other effects were significant; thus, the interference and control groups did not differ in their beliefs about interference effects in paired-associate learning, and span groups did not differ in this regard either. A one-sample t-test revealed that subjects did give lower predictions for the Interference-group (*M* = 35.54, *SD* = 14.60) in comparison to the experimenter-provided control group's performance of 60%, t (124) = -18.74.

Because subjects did not appear to have prior beliefs or knowledge about the source of interference in paired associate tasks, direct knowledge can be eliminated as a source of information for subjects' JOLs here. Although subjects gave significantly lower predictions for the interference than control group in the memory questionnaire, it is unlikely that such a prior belief could influence subjects in the second part of the experiment. Because in the memory questionnaire subjects were comparing two conditions, they will not be able to do this in the actual experiment, which is important because Koriat et al. (2004) was only able to elicit theory-based JOLs from their subjects when they were given two or more conditions (e.g., 10 min. vs. 1 week vs. 1 year).

List 2 recall accuracy

Subjects' list 2 immediate and delayed recall accuracy was analyzed with a 2 (Interference-group: interference, control) x 2 (Word pair type: interference, non-interference) x 2 (Recall test: immediate vs. delayed) repeated measures ANCOVA with WM as the covariate. Subjects recalled more words on the immediate recall test (*M* = .44, *SD* = .21) than the delayed recall test (*M* = .41, *SD* = .21), $F(1, 127) = 29.02$, $\eta_p^2 = .19$, and more non-interference items (*M* = .47, *SD* = .24) than interference items (*M* = .38, *SD* = .23), $F(1, 127) = 16.84$, $\eta_p^2 = .12$. More importantly, subjects in the control condition (*M* = .48, *SD* = .20) recalled more items than subjects in the interference condition (*M* = .37, *SD* = .20), $F(1, 127) = 9.11$, $\eta_p^2 = .07$, indicating that the between-groups PI manipulation was successful.

These two main effects were qualified by an interaction between interference condition and Word pair type, indicating an item-level PI effect, $F(1, 127) = 9.74$, $\eta_p^2 = .07$. In contrast to E1, subjects in the interference condition recalled significantly fewer interference items than non-interference items at immediate and delayed recall tests (see Figure 8), $t(64) = -4.931$, $p < .01$ and $t(64) = -5.153$, respectively. Subjects in the control condition recalled the same proportion of interference and non-interference items at both immediate and delayed recall tests, $t(64) = -0.95$, $p > .05$, and $t(64) = -0.31$, $p > .05$ respectively. WM was not a significant covariate, indicating no differences between high and low WM subjects in interference susceptibility, $F < 1$, $p = .36$. No other effects were significant (p's > .13).

Metacognitive Judgments at Encoding

A 2 (Interference-group: interference, control) x 2 (Word pair type: interference, non-interference) repeated measures ANCOVA with WM as the covariate was conducted on mean JOL magnitude for list 2. Subjects in the control condition gave significantly higher JOLs than did subjects in the interference condition (see Figure 9), $F(1, 126) = 7.40$, $\eta_p^2 = .06$, which is consistent with E1. Despite the fact that item-level PI effects were significant in recall here, subjects in the interference condition did not give significantly higher JOLs to the interference items compared to the non-interference items, $F < 1$, $p = .58$. No other effects were significant.

Finally, a 2 (List: list 1, list 2) x 2 (Word pair type: interference, non-interference) repeated measures ANCOVA with WM as the covariate indicated that subjects in the interference-group gave higher JOLs on list 1 ($M = 42.22$, $SD = 21.78$) than in list 2 ($M =$

34.77, SD = 18.98), $F(1, 63) = 44.96$, $\eta_p^2 = .42$. No other theoretically meaningful effects were significant.

Gamma correlations. As in E1, gamma correlations were calculated between list 2 JOLs and immediate and delayed recall of list 2. A 2 (Word pair type: interference, non-interference) x 2 (recall test: immediate, delayed) x 2 (Interference-group: interference, control) repeated measures ANCOVA with WM as the covariate was conducted. Control (M = .29, SD = .40) and interference (M = .41, SD = .40) subjects' judgments did not differ in their relative accuracy in discriminating between recalled and not recalled items (see Figure 10), $F(1, 95) = 2.01$, $p = .22$. Control and interference subjects' JOLs were equally accurate for interference and non-interference items, $F < 1$, $p = .54$.

WM was a marginally significant covariate, $F(1, 95) = 3.73$, $p = .056$, $\eta_p^2 = .04$. However, the effect was not in the expected direction. To illustrate the WM effect, subjects were grouped into high, medium, and low span groups by conducting a tertiary split on the composite WM scores. The analyses were re-ran using an ANOVA with WM as a second between-subjects variable and Tukey's as the post hoc test. Low spans were more accurate than medium spans ($p = .03$). High spans were equally accurate compared to low spans ($p = .16$) and medium spans ($p = .63$) (see Figure 11). WM interacted with Word pair type, and test, $F(1, 95) = 5.68$, $\eta_p^2 = .06$. To follow-up this 3-way interaction, separate 2 (Word pair type: interference, non-interference) x 2 (test: immediate, delayed) repeated measures ANOVAs were conducted for each span group. No significant effects were found for low and medium spans. However, high span subjects' JOLs better discriminated between recalled and not recalled list 2 items at delayed recall than on immediate recall (see Figure 12), F (1, 37) = 5.83, η_p^2 = .14. This main effect of test for high spans was qualified by an interaction between Word pair type and test, F (1, 37) = 4.45, η_p^2 = .11. High spans' JOLs for interference items better discriminated between recalled and not recalled items for the delayed recall test (M = .41, SD = .52) than the immediate recall test (M = .23, SD = .58), t (39) = -2.947, but their JOLs for non-interference items did not differ in discriminating between items for immediate (M = .34, SD = .54) and delayed (M = .35, SD = .42) tests, t (40) = -.290, $p > .05$.

In summary, subjects were sensitive to interference they were experiencing indicated by controls giving higher JOLs than did interference subjects, who in turn, gave higher JOLs on list 1 than they did on list 2. Subjects were not able to locate, however, the source of the interference, as indicated by the equivalent JOLs that interference subjects assigned to interference and non-interference items. Despite their lack of sensitivity to item level negative transfer, subjects were better able to predict their recall performance in E2 than in E1, where their gamma correlations between their JOLs and recall accuracy were either negative or close to zero. Surprisingly, low spans' JOLs were more accurate than medium spans' JOLs, but they were not significantly more accurate than high spans. High spans' JOLs were more accurate for delayed recall than immediate recall, which could be due to how the JOL was framed (i.e. "If 1 hour from now..."). High spans might have better taken the amount of time elapsed into account when they made their JOLs.

Metacognitive Judgments at Retrieval

A 2 (Word pair type: interference, non-interference) x 2 (DPOK prompt: 1, 2) x 2 (Interference-group: interference, control) repeated measures ANCOVA with WM as a covariate was conducted on the mean magnitude of DPOK judgments. WM was not a significant covariate, indicating that the span groups were equally not sensitive to PI at recall, $F < 1$, $p = .49$. The only significant effect was that subjects gave higher DPOK judgments at the second prompt ($M = 50.24$, SD = 12.43) than at the first prompt ($M = 45.35$, $SD = 12.43$), $F\,(1, 127) = 118.52$, $\eta_p^2 = .48$.

Gamma correlations. I analyzed the gamma correlation data with a 2 (Word pair type: interference, non-interference) x 2 (DPOK prompt: 1, 2) x 2 (Interference-group: interference, control) repeated measures ANCOVA with WM as the covariate. Again, WM was not a significant covariate, indicating that the span groups' DPOKs were equally accurate, $F < 1$, $p = 1.00$. Subjects' relative accuracy increased across DPOK prompts (see Figure 12), $F\,(1, 95) = 4.54$, $\eta_p^2 = .05$. Overall, DPOKs better discriminated between the recalled and not recalled non-interference items than interference items, $F\,(1, 95) = 9.84$, $\eta_p^2 = .09$. Moreover, control subjects' DPOKs were more accurate than interference subjects' DPOKs in discriminating between recalled and not recalled items, $F(1, 95) = 4.19$, $\eta_p^2 = .04$.

Also of importance, interference and control subjects significantly differed in their relative accuracy across the Word pair type variable (see Figure 12), $F\,(1, 95) = 7.54$, $\eta_p^2 = .07$. Interference subjects' DPOKs were more accurate in discriminating between recalled and not recalled non-interference items than interference items for first

DPOK prompt, t (41) = -3.68, $p < .01$ and the second DPOK prompt, t (42) = -2.36. Control subjects' DPOKs were equally accurate in discriminating between recalled and not recalled non-interference and interference items at the first DPOK prompt, t (57) = -.17, $p > .05$ and the second DPOK prompt, t (56) = 1.10, $p > .05$ respectively. No other effects were significant.

Discussion

The primary goal of E2 was to clarify and replicate E1, and the secondary goal was to begin investigating how WM influences monitoring for negative transfer and PI. Here, PI effects were significant at the list and item-level.

Negative transfer. Replicating E1, control subjects gave higher JOLs to list 2 items than interference subjects suggesting that interference subjects were sensitive to negative transfer. Here, however, the difference between controls and interference subjects cannot be attributed to interference subjects taking the immediate recall test of list 1 because there was none. In addition, control and interference subjects' JOLs did equally well in discriminating recalled and non-recalled items, suggesting that the presence of negative transfer did not impair the validity of using it as a basis for their JOLs. Also, replicating E1, interference subjects gave equivalent JOLs to interference and non-interference items despite recalling non-interference items better than interference items. Consequently, interference subjects' sensitivity to negative transfer might be limited to list level awareness of negative transfer. If people are unable to pinpoint which items are causing them difficulty at encoding despite making monitoring judgments, this has different implications for inhibition. Research shows that subjects can delete information from WM (e.g., Hasher, Lustig, & Zacks, 2007). But, if subjects cannot monitor for negative transfer at the item level, then subjects' ability to delete information at the item level must be based on a different cue such as whether the item was relevant or not.

Proactive interference. In contrast to E1, controls and interference subjects did not differ in their mean DPOKs. And, because interference subjects recalled the non-interference items better than interference items, the equivalence of their judgments resulted in controls being more accurate than interference subjects. For the same reason, interference subjects were more accurate discriminating recalled and not recalled non-interference items than interference items. Interference subjects were not sensitive to PI at the list or item levels. Here, interference subjects seemed to be unaware of the competition between items in predicting recall even though they were actually trying to recall the items at the time. Because half the cues were familiar and, in contrast to E1, the words were familiar, English nouns, interference subjects might have felt more confident that the item was about to be recalled. Consequently, familiarity as a basis for metacognitive judgments might be difficult to discount even after 10 or 15 s of attempted retrieval.

The secondary goal of E2 was to begin investigating how WM influences monitoring. I predicted that high spans would be better able to monitor for negative transfer and PI than low spans. Inconsistent with this prediction, low spans' JOLs were more accurate than medium spans, and low and high spans' JOLs were equally accurate. High and low spans' equivalence in monitoring suggests low spans' greater PI susceptibility is due to control failure and not monitoring failure. However, if the function of monitoring is recruiting control, then people, in particular high spans, might not monitor unless there is an opportunity for control. Granted, within 4 s, subjects should have been able to implement some kind of strategy. But, it might be that people behave more strategically in situations where they control the pace of study compared to situations where they do not. Thus, the secondary goal of Experiment 3(E3) was to investigate whether WM actually does not influence monitoring by allowing subjects to control study time.

CHAPTER IV

EXPERIMENT 3

E2 results suggested that subjects were sensitive to PI at the list-level at encoding but were not able to pinpoint a more specific source of the PI. However, subjects' inability to locate the source of PI might be restricted to explicit judgments rather than an actual inability to detect the source. That is, subjects might be able to implicitly monitor for PI at the item level. To test this idea, the same procedures were used in E3 as in E2, with one exception. In E3, a more implicit measure of encoding fluency was used by having subjects control their study time for each item (Koriat & Ma'ayan, 2005) rather than having them make JOLs. Koriat & Ma'ayan (2005) argued that when subjects pace themselves during encoding, the amount of time spent studying an item is determined by the ease of learning the item and that self-paced study time can be used as an index of encoding fluency.

Also, in E3, I further investigated whether individual differences in WM predicted monitoring ability. Subjects might need more time to effectively implement a strategy or other control process, and allowing subjects to control study time provides this opportunity. If monitoring recruits a control process to combat PI, and individual differences in WM are present in controlling PI (Kane & Engle, 2000), then individual differences in WM might predict monitoring ability. High spans should spend more time viewing the word pair than low spans. This prediction was based on Turley-Ames and

Whitfield's (2003) finding that, when subjects are allowed to allocate their time to math problems and to-be-remembered words as they wish, high spans spend more time viewing the word than do low spans. Moreover, when subjects are trained to use a strategy all of Turley-Ames and Whitfield's subjects spent more time viewing the word. If self-paced study time is an index of encoding fluency (Koriat & Ma'ayan, 2005), then the extra time high spans spent studying the word could have been because they were sensitive to the negative transfer and were trying to exert control by using a strategy. If high and low spans do not differ in the time spent viewing the word pair and high spans recall more words, then the relationship between monitoring and control might not follow the predicted pattern of monitoring recruiting control. If low spans spend more time viewing the word pair than high spans, then this would suggest that low spans are aware that they are not encoding the word pairs well enough.

Finally, the contradictory DPOK results from E1 and E2 need to be resolved. E3 used the same procedures as in E2. If the DPOK results from E2 are replicated, then it can be concluded that subjects are not taking interference into account when making their judgments at retrieval.

E3 Method

Subjects

One hundred seven subjects were recruited from the introductory psychology pool at UNCG. All subjects were screened via the same WM tasks as in E2.

Design

Interference condition was a between-subjects variable. Word pair type, List, and DPOK prompt were within-subjects variables, and WM was the covariate.

Materials and Procedure

The same procedure and materials were used as in E2, except subjects controlled their study time and did not make immediate JOLs. Thus, during list learning, subjects advanced from word pair to word pair by pressing the ENTER key on the keyboard. Subjects were instructed to study the word pairs so that when they were presented with the first word they would remember the second word. Subjects were aware that they would have only one opportunity to study the word pair, and that they should only advance to the next word pair when they felt that they had mastered the present word pair. However, subjects were instructed not to spend more time then they needed on any given word pair. Word pairs stayed on screen for at least 3 s until subjects pressed the ENTER key or until 30 s had elapsed. Interference word pairs in list 2 were re-combinations of word pairs from list 1. Non-interference word pairs in list 2 were new cue-response word pairs.

Results

List 2 recall accuracy

For immediate and delayed recall, two separate 2 (Interference-group: interference, control) x 2 (Word pair type: interference, non-interference) repeated measures ANCOVAs with WM as the covariate were conducted. WM did not influence recall accuracy in the same way for the interference and control groups, violating the assumption of homogeneity of covariate regression coefficients, F (1, 103) = 6.31, η_p^2 = .058. The problem with violating this assumption is that the ANCOVA becomes more conservative the more the assumption is violated, resulting in an underestimation of the relationship between the dependent and independent variables and an increase in the probability of a Type II error. The recall data was re-analyzed using stepwise regression analyses (Levine, personal communication, May 23, 2007). Difference scores were calculated for immediate and delayed recall by subtracting the mean for the interference items from the mean for the non-interference items; thus, a positive number indicated more PI. Interference-group, WM, and the interaction between Interference-group and WM were the predictors, and item-level PI was the criterion variable.

For immediate recall, interference-group was a significant predictor of item-level PI, (R^2 = .18). Controls recalled more words than interference subjects (see Figure 13). Moreover, interference subjects recalled a greater percentage of non-interference items than interference items whereas controls did not differ between their recall of non-interference and interference items (see Figure 13). As in E2, WM was not a significant predictor of item-level PI (p = .31, ΔR^2 = .01) nor was the interaction between WM and interference-group significant (p = .15, ΔR^2 = .02).

At delayed recall, interference-group was a significant predictor of item-level PI, (R^2 = .20). Interference subjects recalled a greater percentage of non-interference items than interference items while controls recalled both types of items equally well (see Figure 13). Again, WM was not a significant predictor of item-level PI (p = .10, ΔR^2 =

.02), nor was the interaction between WM and interference-group significant (p = .07, ΔR^2 = .03).

List 2 Study Time

A 2 (Interference-group: interference, control) x 2 (Word pair type: interference, non-interference) repeated measures ANCOVA with WM as the covariate was conducted on the amount of time subjects spent studying the word pairs in list 2 (see Figure 14 & 15). No significant effects were found. Interference and control subjects spent equal time studying word pairs (see Figure 14), F (1, 103) = 1.28, p = .26. Moreover, interference subjects studied the interference and non-interference items equally, F < 1, p = .88. WM also was not a significant covariate, F (1, 103) = 2.51, p = .12 (see Figure 15).

Because interference subjects might have differed in the amount of time they spent studying list 1 versus list 2, a 2 (List: 1, 2) x 2 (Word pair type: interference, non-interference) repeated measures ANCOVA with WM as the covariate was conducted. Again, there were no significant effects. Interference subjects studied list 1 and list 2 for the same amount of time, F (1, 52) = 1.64, p = .21 (see Figure 16). In the ANCOVA, WM was not a significant covariate, F < 1, p = .91 (see Figure 17). However, when WM was correlated with study time separately for control and interference subjects, WM significantly correlated with study time (r = .30) for controls but not for interference subjects (r = .03).

Gamma correlations between study time and recall

A 2 (Word pair type: interference, non-interference) x 2 (Interference-group: interference, control) repeated measures ANCOVA with WM as a covariate was conducted on the list 2 immediate and delayed recall data. For both analyses, no significant effects were found. Interference and control subjects were equally unable to discriminate via study time between recalled and not recalled items for both interference and non-interference items, F (1, 67) = 1.62, p = .21, and F (1, 61) < 1, p = .57, for immediate and delayed recall respectively (see Figure 18). WM was not a significant covariate in either analyses, F < 1, p = .90, and F < 1, p = .33 for immediate and delayed recall respectively.

DPOK judgments

A 2 (Interference-group: interference, control) x 2 (Word pair type: interference, non-interference) x 2 (DPOK prompt: 1, 2) repeated measures ANCOVA with WM as the covariate was conducted on subjects' mean magnitude for their DPOK judgments. Because WM again interacted with the interference-group, the data was re-analyzed using stepwise regression. For each DPOK prompt, difference scores were calculated by subtracting the mean DPOK judgment for interference items from the mean DPOK judgment for non-interference items; thus, a positive number indicates more PI sensitivity at the item-level. Interference-group, WM, and the interaction between interference-group and WM were entered in as predictors, and item-level PI sensitivity as measured by the difference scores for the DPOK prompts was the criterion variable.

For the first DPOK prompt, interference-group was a significant predictor of whether subjects were sensitive to item-level PI, (R^2 = .08). Controls gave higher DPOKs than did interference subjects. Moreover, interference subjects gave lower judgments to the interference items than the non-interference items, whereas controls gave equal judgments to both types of items (see Figure 19). WM was not a significant predictor item-level PI sensitivity (p = .74, ΔR^2 = .00), but the interaction between WM and interference-group was significant (ΔR^2 = .04). For the second DPOK prompt, interference-group was not a significant predictor of whether subjects were sensitive to item-level PI, (R^2 = .01, p = .34), nor was WM (ΔR^2 = .00, p = .63), but the interaction between WM and interference-group was again significant (ΔR^2 = .05).

To follow-up the interaction between WM, interference-group, and Word pair type, three separate 2 (Interference-group: interference, control) x 2 (DPOK prompt: 1, 2) x 2 (Word pair type: interference, non-interference) ANOVAs were conducted on mean DPOK magnitude for each span group (see Figure 20). Low spans gave higher judgments to non-interference items than they did for interference items, $F(1, 33) = 4.40$, $\eta_p^2 = .12$, but they did this in both the control and interference groups, $F(1, 33) = 1.00$, $p = .33$. In contrast, medium and high spans' DPOKs did differ across Word pair type and interference-group, $F(1,34) = 4.81$, $\eta_p^2 = .12$ and $F(1,33) = 4.18$, $\eta_p^2 = .11$, respectively. Medium and high spans in the interference group gave higher judgments to the non-interference items than they did to the interference items, $t(19) = -2.65$ and $t(16) = -2.14$, respectively. In contrast, medium and high spans in the control condition gave equivalent judgments to the interference and non-interference items, $t(15) < 1$, $p = .60$ and $t(17) < 1$, $p = .65$, respectively.

Gamma correlations between DPOK judgments and delayed recall

A 2 (Interference-group: interference, control) x 2 (Word pair type: interference, non-interference) x 2 (DPOK prompt: 1, 2) repeated measures ANCOVA with WM as the covariate was conducted on subjects' gamma correlations for their DPOK judgments. WM was not a significant covariate in either analysis, F (1, 55) = 1.30, $p > .05$, η_p^2 = .023. Subjects were more accurate on the second DPOK prompt than the first DPOK prompt, F (1, 55) = 5.31, η_p^2 = .09. Interference subjects' DPOKs were less accurate for the interference items than the non-interference items; controls were equally accurate for both types of items (see Figure 21), F (1, 55) = 7.09, η_p^2 = .11.

Discussion

The primary goal of E3 was to investigate whether subjects' were sensitive to negative transfer and PI at the item-level if they made implicit judgments (controlled study time) rather than explicit judgments (JOLs) at encoding. The secondary goal was to investigate whether WM actually did not influence monitoring.

Negative transfer. Consistent with the discrepancy-reduction model (e.g., Nelson & Narens, 1990; see Kornell & Metcalfe (2006) for an opposing view) which predicts that subjects study difficult items longer than easier items, I hypothesized that people would spend more time studying interference items than non-interference items, thereby indicating sensitivity to PI. Inconsistent with these predictions, but consistent with the first two experiments, E3 suggested that subjects were not sensitive to negative transfer at the item level: interference subjects studied the interference and non-interference word pairs equally. But, here, control and interference subjects also did not differ in the amount of time spent studying the word pairs, in contrast to E1 and E2. Here, then, subjects seemed unaware of negative transfer at the list level. Furthermore, interference subjects did not spend more time studying list 2 than list 1, which is also inconsistent with E1 and

E2, where subjects gave lower JOLs to list 2 than list 1. E3's inconsistencies have two possible sources: how study time was conceptualized, and learning-to-learn's influence on encoding fluency.

The equal time that subjects spent studying the interference and non-interference word pairs could have been due to the opposing influences of negative transfer and learning-to-learn. Postman's (1972) discussion of transfer effects cites Ward (1937), who had subjects learn 16 lists of nonsense syllables. By the last list, subjects reached the criterion level of learning twice as fast as they did on the first list. In the present experiment, I was not measuring how quickly subjects learned a list to a specific criterion, but if subjects tend to be faster in *actual* learning, then this acceleration in learning a list could have influenced subjects' perceived ease of learning the second list. For instance, without negative transfer and given the opportunity to control study time, subjects might spend less time studying subsequent lists in a multi-list learning experiment because they have become accustomed to the experimental procedure. In the presence of negative transfer, the contribution of learning-to-learn would still be present, but it could be opposed by the influence of negative transfer. Thus, if subjects are sensitive to list-level or item-level negative transfer and consequently spend more time studying items than they would have if negative transfer were not present, then these two processes might cancel each other out. Consistent with this idea, interference subjects spent equal time studying list 1 and 2, and they studied the word pairs almost twice as long as subjects did in the first two experiments where presentation rate was fixed. To test whether the above speculation is correct, a follow-up experiment should include a more appropriate control group that studied multiple unrelated lists to minimize negative transfer and equate learning-to-learn effects on encoding fluency.

Proactive Interference. Consistent with E1, but inconsistent with E2, interference subjects gave lower DPOK judgments than controls, suggesting that interference subjects were sensitive to list-level PI. In addition, interference subjects gave lower DPOK judgments to the interference items than the non-interference items, suggesting that they were sensitive to item-level PI. Overall, DPOKs were very accurate in predicting eventual recall; however, interference subjects' judgments were more accurate for interference items than non-interference items. In comparison, controls were equally accurate for both types of items. It appears, then, that interference subjects were still misled to some degree by the familiarity of the interference word pairs. Why did E3 replicate E1 whereas E2 did not? I address this perplexing question in the general discussion.

Regarding WM, high spans recalled more words than medium spans who recalled more words than low spans. However, in the interference condition, high and low spans recalled relatively the same percentage of words while medium spans recalled the smallest percentage of words. The span groups studied the interference and non-interference word pairs for the same amount of time regardless of whether they were in the control or interference condition, suggesting they were not sensitive to negative transfer at the item-level or list-level. Because span groups studied the word pairs for an equal amount of time, the possible opposing influences of learning-to-learn and negative transfer appears to have influenced all subjects equally. Consistent with their recall, medium spans gave lower DPOK judgments than high and low spans in the interference condition. Furthermore, high, medium, and low spans were equally accurate in discriminating whether they were about to recall a word or not. Taken together, the results suggest that subjects, regardless of working memory ability, are sensitive to PI at recall, but what they can do about it is less certain. The atypical recall results (as in E2) suggest that high and medium span subjects were not engaging control processes despite being able to monitor for PI.

CHAPTER V

GENERAL DISCUSSION

The motivation for the 3 experiments reported here was to investigate how people *know* to exert control to combat negative transfer and PI. Consistent with the dominant view of monitoring and control in the metacognitive literature (e.g., Nelson & Narens, 1990), the negative transfer/PI sensitivity hypothesis proposed that people monitor for the experience of negative transfer or PI. If they detect negative transfer or PI, they exert control processes to resolve it. Thus, the ability to overcome PI relies on accurate monitoring. The present research focused on three primary questions: 1) Are people sensitive to negative transfer at encoding? 2) Are people sensitive to PI at retrieval? 3) If so, does sensitivity to negative transfer and PI vary with working memory ability? Below I address each question by first summarizing the results and then discussing the theoretical implications for monitoring and control. Because individual differences in WM have been shown to influence PI susceptibility (Kane & Engle, 2000; Rosen & Engle, 1998), WM was included in the analyses to help clarify the results. Thus, I addressed the 3rd question within the two larger sections: sensitivity to negative transfer and PI.

Are people sensitive to negative transfer at encoding?

As mentioned previously, few studies (Diaz & Benjamin, 2005; Eakin, 2005; Maki, 1999; McGuire & Maki, 2001; Metcalfe et al., 1993; Schreiber, 1998; Schreiber &

Nelson, 1998) have investigated whether subjects use the presence of interference as a basis for metacognitive judgments. Only two of these studies (Diaz & Benjamin, 2005; Metcalfe et al., 1993) have induced PI in their subjects, and only three had subjects make judgments prior to any recall attempts (Diaz & Benjamin, 2005; Schreiber, 1998; Schreiber & Nelson, 1998). Consequently, very little research has been done on whether people might monitor for negative transfer at encoding. Part of the contribution of the present research was to begin investigating this question.

In short, subjects seemed to be sensitive to negative transfer at the list level but not at the item level, at least when making explicit judgments. Consistent with their recall performance, controls gave higher JOLs than did interference subjects, and interference subjects gave higher JOLs to list 1 than list 2. In E2, this pattern was found even after the immediate recall test of list 1 was eliminated, indicating that interference subjects' lower JOLs were not due to the experience of retrieval fluency during list 1 recall. Inconsistent with their recall, however, interference subjects gave interference and non-interference items equivalent JOLs and study time, indicating they were unable to locate a more specific source of negative transfer than the entire list. Furthermore, subjects' list-level sensitivity to negative transfer appeared to be restricted to explicit judgments, as indicated by the equal amount of time spent by controls and interference subjects when they were allowed to pace themselves during study in E3. The results were inconclusive as to whether the presence of negative transfer impaired the relative accuracy of subjects' JOLs: In E1 and E3, subjects were unable to discriminate between items that would later be recalled versus not recalled, but in E2, subjects were able to do so.

The present experiments suggest that subjects were sensitive to negative transfer at the list level but not at the item level. If people are restricted to only a list-level awareness, this could restrict how they initiate control processes. For inhibition, the inability to locate a more specific source of negative transfer suggests that all irrelevant information, regardless of whether it is causing interference, needs to be suppressed. Although consistent with previous research (e.g, Bjork, 1970; Postman et al., 1968), suppression comes with a later retrieval cost for items that are suppressed (Rosen & Engle, 1998). Consequently, it would be more advantageous to be selective in targeting what is suppressed. For more overt strategies, such as rehearsal, an item-level awareness of negative transfer might be less necessary for instituting appropriate control changes. If overall difficulty in encoding is detected, then this should be sufficient in prompting subjects to switch to a better strategy (see Sahakyan et al., 2004).

Hasher, Zacks, and their colleagues (Hasher et al., 2007; Hasher, Zacks & May, 1999) proposed 3 functions for inhibitory control: *access*, *restraint*, and *deletion*. Of the three, *access* and *deletion* are the most relevant here. *Access* referred to people's ability to keep irrelevant information out of the focus of attention before it intrudes. *Deletion* referred to people's ability to remove irrelevant information from the focus of attention should it intrude, or when it is no longer needed. If people are limited by having only a list-level awareness of negative transfer, then they can only inhibit identified competitors. If this is the case, then the most advantageous strategy might be to exert control before negative transfer ever happens (i.e., using the *access* function of inhibition). For example, in Kane and Engle (2000), subjects completed a recall test after each list. The recall test might have served as an implicit cue that the prior-list information was now irrelevant and should be suppressed. In contrast, in the present E2 and E3, the immediate recall test of list 1 was eliminated (and no instruction to disregard list 1 was given), and the typical recall superiority of high spans over low spans was not found.

As previously mentioned, a list-level awareness of negative transfer is consistent with the ostensible role of strategy change in producing directed forgetting effects. Sahakyan et al. (2004) found that making a global JOL for list 1 had equivalent recall benefits for list 2 as did a forget cue for list 1. Sahakyan et al. argued that the normal tendency for subjects told to remember list 1 is to *not* evaluate the efficacy of their list 1 strategy. In the present experiments, subjects only were sensitive to list-level negative transfer if they made explicit judgments like JOLs. When they were allowed to control study time and their judgments were thus implicit, controls and interference subjects did not differ in study time, suggesting they engaged in their normal habit of not evaluating their learning.

However, were subjects really insensitive to item-level negative transfer and do subjects really not evaluate at the list level if they are not explicitly prompted to do so? E3 subjects spent almost twice as long as E1 and E2 subjects in trying to memorize the word pairs. So, even if subjects were not evaluating the efficacy of their strategy (e.g., staring at the word, rehearsal, making a story), it seems unlikely they would spend so long trying to memorize the word pairs if they did not feel they needed to study them longer. However, the reasoning behind subjects' behavior might not be interference because there were no differences in study time between control and interference subjects. But, subjects in multi-list, non-interference situations might spend less time studying list 2. If this is the case, then the lack of a difference between interference subjects' list 1 and list 2 study times might indicate a sensitivity to list-level negative transfer. Likewise, the lack of a difference between interference subjects' study time for interference and non-interference items might indicate a sensitivity to item-level negative transfer. Future studies should include both subject-controlled study time and JOLs with appropriate control groups (two-list control group with low-interference items to assess the effect of learning-to-learn) to further assess whether subjects are sensitive to negative transfer via implicit judgments.

In addition to the effect of learning-to-learn, people's beliefs about interference is potentially important in investigating whether people monitor for negative transfer and PI. In E2, subjects predicted that other people in the interference condition would recall fewer words than the people in the control group (experimenter provided control group estimate). Thus, people might have a belief that the more items a person learns the harder it is to recall information. The present research could then be explained by an interference-belief explanation rather than a sensitivity to interference explanation. However, as with Koriat et al. (2004), subjects might only be able to correctly predict recall when they can compare different conditions. What was more intriguing was that subjects predicted that interference items would be better recalled than non-interference items, which bolsters the sensitivity to negative transfer/PI explanation. In essence, despite their beliefs, interference subjects gave lower JOLs than did controls. Furthermore, this raises the question of how their mistaken belief about interference items might have influenced the allocation of study time in E3. I did not assess subjects' beliefs about interference in E3. Thus if E3 interference subjects noticed that certain word pairs consisted of repeated cues and responses, they might have mistakenly believed they would be easier to remember and studied them less than they would have. Moreover, interference subjects might have given higher JOLs to interference items than they would have if they did not have such a belief. Furthermore, the JOLs for interference items might have been even higher if not for subjects' sensitivity to negative transfer.

Working memory. Did sensitivity to negative transfer vary with WM? In E2, WM did not significantly influence recall performance or JOLs. High spans' JOLs were more accurate in discriminating pairs that would be recalled versus would not be recalled at the delayed test than at the immediate test. In E3, WM did not influence recall performance, study time, or the relative accuracy of implicit monitoring. Thus, span groups did not differ in their monitoring ability. This could suggest that the difference between span groups might be their ability to control negative transfer and not in their ability to monitor for it. This suggestion might seem odd after considering that in the present research span groups did not differ in PI susceptibility at recall. But, as mentioned previously, high and medium spans might not have been suppressing list 1 as they normally would have if they had completed a list-1 recall test. WM might still influence monitoring ability. Future research could use a forget instruction after list 1 to prompt high spans to use suppression.

Are people sensitive to PI at retrieval?

The results for retrieval were more difficult in interpreting than the results for encoding. First, all 3 experiments had contradictory results as to whether interference subjects were sensitive to PI at the item level, despite the 3 experiments using the same procedure for assessing PI sensitivity. Second, for list-level sensitivity to PI, E2 contradicted E1 and E3 results. I first summarize the results, and then I discuss how the present results coincide with previous research.

E1 and E3 suggest that subjects were sensitive to PI at the list level: Consistent with recall, control subjects gave higher DPOK judgments than did interference subjects. In contrast, in E2, control and interference subjects gave equivalent DPOK judgments, suggesting they were not sensitive to list-level PI. At the item-level, the picture was more complicated: In E1, interference subjects did not differ in their recall of interference and non-interference items. However, in E1, interference subjects gave slightly higher DPOK judgments to interference items than they did to non-interference items (p = .07), indicating they were being misled by cue familiarity. In contrast, in E3, consistent with their recall, interference subjects gave lower DPOKs to interference items than they did to non-interference items, indicating item-level sensitivity to PI. In E2, interference subjects gave equivalent DPOKs to interference and non-interference items, indicating they were not sensitive to PI at the item level but were not misled by cue familiarity either. Controls gave equivalent judgments to the two types of items in all 3 experiments.

In all 3 experiments, subjects' judgments were very accurate overall. However, also in all 3 experiments, interference subjects' DPOKs for interference items were less accurate than their DPOKs for non-interference items. However, interference subjects' lowered accuracy for interference items over non-interference items could be attributed to different sources: cue familiarity in E1, PI in E3, and equivalence in DPOK magnitude in E2. In comparison, in all 3 experiments, controls were equally accurate for both interference and non-interference items.

The item-level effect in E3 is consistent with Maki (1999). Her subjects learned number-word pairs via study/test cycles during the learning phase of her experiment. That is, subjects gained experience in which number-word pairs they had trouble recalling. Subjects needed to reach a criterion of 6 out of 12 number-word pairs correct on list 1 before moving on to list 2. Maki found that subjects who learned interference number-word pairs in list 2 gave lower JOLs to list 1 than did subjects who learned non-interference number-word pairs in list 2.

But, item-level sensitivity to PI in E3 is inconsistent with Metcalfe et al. (1993), who found that feeling-of-knowing judgments for word pairs that subjects did not correctly recall were higher for interference word pairs than for non-interference word pairs. However, Metcalfe et al.'s results are consistent with the present E1 where interference subjects gave slightly higher DPOKs to interference items than they did to non-interference items. The item-level PI effect in E3 also is inconsistent with the lack of an item-level effect in E2.

How might these contradictory results be reconciled? It could be that item-level sensitivity to interference is difficult to elicit because at the item level other available cues might seem more diagnostic than interference. For instance, because cue familiarity prompts people to continue to search for the item, cue familiarity's influence might be difficult to discount when making DPOKs. Moreover, when people are encouraged to make quick judgments, as they were here, they are more likely to base their metacognitive judgments on familiarity (Benjamin, 2005). Also, people might base their judgments at retrieval on interference at the item level only after studying and testing themselves on the material repeatedly as in Maki's study. Or, they might only detect it at the item level if there is a high level of competition between items as there would be with re-combined word pairs (E3). Why, then, weren't subjects sensitive to PI at the item-level in E2? On one hand, E2's results could be due to chance. On the other hand, the types of control strategies (e.g., inhibition, rehearsal, etc.) elicited by list-level monitoring at encoding might have differentially influenced how confident subjects were that they were about to retrieve an item at recall. Subjects in E1 and E3 might have approached the task differently than subjects in E2: In E1, because the word pairs were foreign language vocabulary, subjects might have been more inclined to use strategies. Of course, this is pure speculation. In E3, subjects spent almost twice as long studying the word pairs, which could have been because they were attempting to use the same strategy for all word pairs. This would be consistent with subjects' behavior in the OSPAN. When subjects are given a strategy, they spend more time studying the words than when they are not given a strategy (Turley-Ames & Whitfield, 2003). Despite the lack of an item-level study time effect, E3 subjects still could have been aware that learning these word pairs were difficult. And, when they actually had trouble recalling the interference word pairs, their earlier observation might have been validated. Again, this is speculation and research will need to be done to test these ideas.

In contrast, list-level awareness of PI seemed to be easier to elicit than item-level awareness. In E1 and E3, subjects were sensitive to PI at the list level, suggesting that subjects should exert control prior to retrieval. People might have developed a theory that focusing on the present material is best for recall. Again, the *access* function of inhibition would be the most useful since subjects could suppress an entire list, and it would not require identification of specific competitors.

Working memory. WM was included in E2 and E3 in order to help clarify results. If monitoring recruits control and WM influences PI susceptibility, then monitoring for PI might vary with WM. In E2, WM did not predict subjects' DPOKs or influence the accuracy of their DPOKs. In contrast, in E3, WM did influence subjects' DPOKs: Low spans' gave higher DPOKs to non-interference items than they did to interference items. They did this in both the control and interference conditions, suggesting the difference might have been due to intrinsic qualities (e.g., concreteness, distinctiveness) about the non-interference word pairs rather than competition between the re-combined interference word pairs. High and medium spans also gave higher DPOKs to the non-interference items than the interference items, but only if they were in the interference condition, suggesting that they were sensitive to PI at the item level whereas low spans were not WMC did not influence the relative accuracy of subjects' DPOKs, nor did it predict recall performance, indicating that all subjects experienced relatively the same amount of PI.

Experiencing PI might be a normal experience for low spans. In contrast, high spans might be better able to overcome PI when they encounter it as suggested by Kane and Engle's (2000) results where only high spans in the divided attention conditions differed from their controls on PI buildup lists. In contrast, low spans in the divided attention conditions did not differ from the low spans in the control group. The atypical nature of the experience might have aided high and medium spans in recognizing that the interference items were more difficult to recall because it drew their attention to it. But, if high and medium spans were monitoring at the item level, why did they not better exert control during retrieval? High and medium spans might have realized the cost of suppression. They might not have suppressed word pairs because they expected to be tested on list 1 (either because they explicitly asked, were suspicious, or because of how the JOLs were framed). Moreover, the AB/ABr method of inducing interference could prevent the use of certain strategies. In both E2 and E3, suppressing a competitor from list 1 would have aided trying to recall the second word in the current word pair, but it would have harmed recall for a later word pair since the interference word pairs were re-combinations from list 1. This suggests that un-binding and re-binding recently learned information within a short time is difficult for even high spans, especially if they cannot use inhibition to their advantage during the task.

Conclusion

The goal of the present research was to begin investigating whether people are aware of negative transfer and PI, because such monitoring might prompt people to try to overcome interference through control processes. Despite monitoring not translating into control here, the present research does suggest that monitoring might recruit control processes as indicated by the high and medium spans monitoring at the item level during recall. Future research should focus on the coordination of monitoring and control in order to demonstrate that monitoring does indeed recruit control processes. Just as importantly, more investigation into whether or not people are restricted to a list-level awareness of negative transfer at encoding is needed as well. In contexts affording the opportunity for control, item-level awareness of increased negative transfer could translate into a more efficient and effective use of inhibition and better recall.

REFERENCES

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.A. Bower (Ed.). *The psychology of learning and motivation* (vol.8, pp. 47-89). New York: Academic Press.

Baddeley, A.D., & Logie, R.H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). New York: Cambridge University Press.

Begg, I., Duft, S., LaLonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28*, 610-632.

Benjamin, A. S. (2005). Response speeding mediates the contributions of cue familiarity and target retrieveability to metamnemonic judgments. *Psychonomic Bulletin & Review, 12*, 874-879.

Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior*, *9*, 255-268.

Bjork, R. A., LaBerge, D., & Legrand, R. (1968). The modification of short-term memory through instructions to forget. *Psychonomic Science*, *10*, 55-56.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624-652.

Conway, M. A., Harries, K., Noyes, J., Racsma'ny, M., & Frankish, C. R. (2000). The disruption and dissolution of directed forgetting: Inhibitory control of memory. *Journal of Memory and Language*, *43*, 409-430.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12,* 769-786.

Diaz, M. & Benjamin, A. S. (2005, November). Effects of proactive interference and release from PI on judgments of learning. Poster presented at the 46th annual meeting of the Psychonomics Society, Toronto, Canada.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*, 374-380.

Eakin, D. K. (2005). Illusions of knowing: Metamemory and memory under conditions of retroactive interference. *Journal of Memory and Language*, *52,* 526-534.

Ekstrom, R. B., French, J. W., Harman, M. H., & Derman, D. (1976). *Manual for kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Services.

Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse, (Eds.), *Variation in working memory* (pp. 227-249). New York: Oxford University Press.

Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat, (Eds.), *Attention and performance XVII:*

Cognitive regulation of performance: Interaction of theory and application (pp. 653-675). Cambridge, MA: MIT Press.

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging, 17,* 209-225.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 336-358.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*, 643-656.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490-517.

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 34-53.

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language,* 52, 478-492.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of experimental psychology: Learning, Memory, and Cognition, 32,* 609-622.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. GruneBerg P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). London: Academic Press.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, *130*, 199-207.

Maki, R. H. (1999). The roles of competition, target accessibility, and cue familiarity in metamemory for word pairs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1011-1023.

May, C., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. *Memory and Cognition*, *27*, 759-767.

McGuire, M. J., & Maki, R. H. (2001). When knowing more means less: The effect of fan on metamemory judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1172-1179.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 851-861.

Miyake, A. & Shah, P. (Eds.) (1999). Models of working memory: Mechanisms of active maintenance and executive control. New York: Cambridge University Press.

Muther, W. S. (1965). Erasure or partitioning in short-term memory. *Psychonomic Science*, *3*, 429-430.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, *51*, 102-116.

Nelson, T.O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory, 2,* 325-335.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation* (vol. 26, pp. 125-141). San Diego, CA: Academic Press.

Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General, 134,* 368-387.

Postman, L. (1972). Transfer, interference, and forgetting. In J. W. Kling & L. A. Riggs (Eds.), Experimental Psychology: Vol. 2. Learning, motivation, and memory (3rd ed., pp. 1019-1132). New York: Holt, Rinehart, and Winston, Inc.

Postman, L., Stark, K., & Fraser, J. (1968). Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior*, *7*, 672-694.

Radvansky, G. A., & Zacks, R. T. (1991). Mental models and the fan effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 940-953.

Rosen, V. M., & Engle, R.W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General, 126,* 211-227.

Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language*, *39*, 418-436.

Sahakyan, L. (2004). Destructive effects of "forget" instructions. *Psychonomic Bulletin and Review, 11*, 555-559.

Sahakyan, L., & Delaney, P. F. (2003). Can encoding differences explain the benefits of directed forgetting in the list method paradigm? *Journal of Memory and Language*, *48*, 195-206.

Sahakyan, L., Delaney, P. F., & Kelley, C. M. (2004). Self-evaluation as a moderating factor of strategy change in directed forgetting benefits. *Psychonomic Bulletin and Review*, *11*, 131-136.

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1064-1072.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207-217.

Schreiber, T. A. (1998). Effects of target set size on feelings of knowing and cued recall: Implications for the cue effectiveness and partial-retrieval hypotheses. *Memory and Cognition*, *26*, 553-571.

Schreiber, T. A., & Nelson, D. L. (1998). The relation between feelings of knowing and the number of neighboring concepts linked to the test cue. *Memory and Cognition*, *26*, 869-883.

Turley-Ames, K. J., & Thompson, H. M. (2003, November). Working memory, metacognitive awareness, and reading comprehension. Poster presented at the 44th annual meeting of the Psychonomics Society, Vancouver, Canada.

Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language*, *4*9, 446-468.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127-154.

Underwood, B. J. (1945). The effect of successive interpolation on retroactive and proactive inhibition. *Psychological Monographs*, *59*(1, Whole No. 273).

Vernon, D., & Usher, M. (2003). Dynamics of metacognitive judgments: Pre- and postretrieval mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 339-346.

Whitely, P. L. (1927). The dependence of learning and recall upon prior intellectual activities. *Journal of Experimental Psychology, 10*, 448-508.

Appendix A

E1: Four Versions of List 1

List 1 version	1		2		3		4	
Swahili cues	tabibu	python	ladha	banker	tabibu	tomato	ladha	prayer
repeat in List 2	buu	banker	ndoo	python	buu	prayer	ndoo	tomato
	maiti	squire	sumu	squire	maiti	cotton	sumu	cotton
	pipa	cancer	zulia	cancer	pipa	almond	zulia	almond
	yatima	ambush	tumbili	ambush	yatima	oyster	tumbili	oyster
	samadi	tribal	zeituni	tribal	samadi	harbor	zeituni	harbor
	punda	cavity	bustani	cavity	punda	custom	bustani	custom
	handaki	ballad	baharia	ballad	handaki	grapes	baharia	grapes
	kasuku	dipper	kaputula	dipper	kasuku	shroud	kaputula	shroud
	mshoni	ration	jibini	ration	mshoni	quarry	jibini	quarry
Trial 1 Mean Trial 2 Mean Trial 3 Mean	not available		not available		not available		not available	
Swahili cues	nyanya	tomato	nyanya	tomato	nyanya	banker	nyanya	banker
do not repeat	sala	prayer	sala	prayer	sala	python	sala	python
in list 2	pamba	cotton	pamba	cotton	pamba	squire	pamba	squire
	lozi	almond	lozi	almond	lozi	cancer	lozi	cancer
	chaza	oyster	chaza	oyster	chaza	ambush	chaza	ambush
	bandari	harbor	bandari	harbor	bandari	tribal	bandari	tribal
	desturi	custom	desturi	custom	desturi	cavity	desturi	cavity
	zabibu	grapes	zabibu	grapes	zabibu	ballad	zabibu	ballad
	sanda	shroud	sanda	shroud	sanda	dipper	sanda	dipper
	chimbo	quarry	chimbo	quarry	chimbo	ration	chimbo	ration
Trial 1 Mean*	0.11		0.11		not available		not available	
Trial 2 Mean	0.36		0.36					
Trial 3 Mean	0.58		0.58					
* Normative mean likelihood of recall								
across 3 trials, as reported by								
Nelson & Dunlosky (1994)								

Appendix B

E1: List 2 Versions

List 2 version	1		2	
Swahili cues repeated from list 1	tabibu buu maiti pipa yatima samadi punda handaki kasuku mshoni	doctor maggot corpse barrel orphan manure donkey trench parrot tailor	ladha ndoo sumu zulia tumbili zeituni bustani baharia kaputula jibini	flavor bucket poison carpet monkey olives garden sailor shorts cheese
Trial 1 Mean* Trial 2 Mean Trial 3 Mean	0.15 0.47 0.66		0.14 0.43 0.64	
Swahili cues not repeated from list 1	ladha ndoo sumu zulia tumbili zeituni bustani baharia kaputula jibini	flavor bucket poison carpet monkey olives garden sailor shorts cheese	tabibu buu maiti pipa yatima samadi punda handaki kasuku mshoni	doctor maggot corpse barrel orphan manure donkey trench parrot tailor
Trial 1 Mean Trial 2 Mean Trial 3 Mean	0.14 0.43 0.64		0.15 0.47 0.66	
* Normative mean likelihood of recall				
across 3 trials, which were calculated				
based on the normative mean likelihood				
reported by Nelson & Dunlosky (1994)				

Appendix C

Global JOL Information

Subjects made global JOLs as well as specific-item JOLs; however, to reduce the length and the redundancy of my results, I did not include the results of the global JOLs for E1 & 2 in the main text. In making global JOLs, subjects answered the following question:

What percent of the Swahili--English word pairs in this list do you think you would remember if you were presented with only the Swahili word 1 hour from now? Please make your rating on a scale from 00% (definitely will not remember) to 99% (definitely will remember).

After the last list presentation, subjects in the immediate and delayed JOL conditions made a global JOL for the immediate recall test: *What percent of the Swahili--English word pairs in this list do you think you'll remember in the immediately upcoming recall test?* Immediately after the last list 2 presentation, subjects in the interference condition made additional "global" JOLs for specific types of word pairs in the following order: Delayed recall global JOL for interference items, Delayed recall global JOL for non-interference items, Immediate recall global JOL for interference items, and Immediate recall global JOL for non-interference items. Delayed recall global JOL for interference items asked the following question:

As you know, some of the Swahili words on this last list had dual meanings (that is they had a different translation here than on the first list). If we tested your memory 1 hour from now, for only the 2nd meanings of these Swahili words, not the first, what percent of your answers do you think would be correct?

Delayed recall global JOL for non-interference items ask the following question:

Also, in this last list, some of the Swahili items were new ones. That is, you had not seen them in the first list. In 1 hour from now, if we tested your memory for the meanings of these new words, what percent of your answers do you think would be correct?

Immediate recall global JOL for interference items asked the following question:

As you know, some of the Swahili words on this last list had dual meanings (that is they had a different translation here than on the first list). If we tested your memory IMMEDIATELY, for only the 2nd meanings of these Swahili words, not the first, what percent of your answers do you think would be correct?

Immediate recall global JOL for non-interference items asked the following question:

Also, in this last list, some of the Swahili items were new ones. That is, you had not seen them in the first list. In an IMMEDIATE recall test, if we tested your memory for the meanings of these new words, what percent of your answers do you think would be correct?

A 2 (Interference-group: control, interference) x 2 (JOL type: immediate, delay) x 3 (Trial: 1st, 2nd 3rd presentation of list) repeated-measures ANOVA was conducted on the mean magnitude of global JOLs for list 2. Global JOLs were list level (i.e. % correct) judgments made at the end of each of 3 study trials, predicting overall performance on a delayed test. Interference-group and JOL type were between-subjects variables, and Trial was the within-subjects variable. Global JOLs increased significantly across trials, F $(2,154) = 43.65$, $p < .01$, and paired t-tests indicated that subjects gave significantly higher global JOLs at each presentation of list 2 (Ms = 25.16, 32.43, 40.42; SDs = 19.32, 20.09, 22.59, respectively) ($p < .01$ for all 3 comparisons). More importantly, subjects in the control condition gave higher global JOLs ($M = 37.43$, $SD = 18.20$) than did subjects in the Interference-group ($M = 28.38$, $SD = 18.21$) suggesting that interference subjects were sensitive to the PI at the list-level, $F\ (1, 77) = 5.01$, $p < .05$. No other main effects and interactions were significant. Subjects' global JOLs did not differ by the type of JOL made (i.e. immediate vs. delayed).

But, I did find a marginally significant interaction between Interference-group and Trial, $F\ (2, 154) = 2.66$, $p = .07$. Control and interference subjects global JOLs did not differ from each other across the first two learning trials, but they did differ from each other on the 3rd trial (see Figure 2). Control subjects gave significantly higher global JOLs than did interference subjects on the 3rd trial, $t\ (79) = 2.81$, $p < .01$.

In addition to the 3 global JOLs for delayed recall, subjects also made a single global JOL for immediate recall performance. Subjects made the immediate recall global JOL after the final presentation of list 2 (i.e. 3rd presentation for Interference-group and 4th presentation for control condition). A 2 (Interference-group: control, interference) x 2 (JOL type: immediate, delay) ANOVA was conducted on the magnitude for a global JOL for immediate recall of list 2. Consistent with the global JOLs for delayed recall, subjects in the control condition ($M = 41.45$, $SD = 26.00$) gave a higher global JOL for immediate recall than did subjects in the Interference-group ($M = 30.56$, $SD = 19.25$), $F\ (1, 77) = 4.65$, $p < .05$. No other effects were significant. Subjects' immediate recall global JOLs were not influenced by the type of item-specific JOL they had been making (i.e. immediate vs. delayed).

Subjects in the Interference-group also made global JOLs by Word pair type (interference vs. non-interference items). Two of these JOLs asked subjects to make a judgment regarding the immediate recall test of list 2 and 2 of these JOLs asked subjects to make a judgment regarding the delayed recall test of list 2. Thus, a 2 (JOL type: immediate, delayed) x 2 (Word pair type: non-interference, interference) x 2 (Recall Test: immediate, delay) repeated-measures ANOVA was conducted on the magnitude for the global JOLs by pair type. Although I did not find any significant effects, I did find that interference subjects gave slightly higher global JOLs to the interference items (M = 26.07, $SD = 17.14$) than the non-interference items ($M = 23.81$, $SD = 16.70$) regardless of whether the judgment was made for an immediate or delayed recall test, F (1, 40) = 3.73, p = .06. Although this marginal effect must be interpreted with caution, it does suggest that interference subjects might have been influenced by the familiarity of the cues for the interference items because they had seen those items twice.

Furthermore, if interference subjects' global JOLs are sensitive to list-level PI, then their list 2 global JOLs should be lower than their list 1 global JOLs. A 2 (JOL type: delayed, immediate) x 3 (Trial: 1st, 2nd 3rd presentation of list) x 2 (List: 1, 2) repeated-measures ANOVA was conducted on the mean magnitude of the list 1 and 2 global JOLs. Interference subjects gave higher global JOLs on each subsequent trial (*M*s= 25.32, 31.10, 37.5; *SD*s = 17.18, 15.82, 17.56, respectively), F (2, 80) = 26.96, $p < .01$, suggesting that subjects might have felt that the ease of encoding the items was becoming less difficult. In order to conduct follow-up tests for some of the effects, I created a dummy variable by collapsing across another variable. For example, the overall global JOL magnitude for each trial reflects an average of list 1 and 2 global JOLs. The estimated marginal means were used rather than the observed means because the estimated marginal means removes the error present in the observed means, and standard deviations were calculated by multiplying the standard error with the square root of the sample size. Moreover, interference subjects gave lower global JOLs to list 2 (M = 28.38, SD = 15.15) than list 1 (M = 34.66, SD = 17.86), F (1, 40) = 14.85, $p < .01$, suggesting that subjects might have been sensitive to the list-level PI present. I found a marginally significant main effect of JOL type, F (1, 40) = 3.53, p = .07. Interference subjects in the immediate JOL condition (M = 35.94, SD = 15.26) gave higher global JOLs than did interference subjects in the delayed JOL condition (M = 27.09, SD = 15.26). No other effects were significant.

Interference subjects list 1 and list 2 global JOLs for immediate recall were also compared (these were single metacognitive judgment made after the 4th and 3rd presentation of list 1 and 2 respectively). A 2 (List: 1, 2) x 2 (JOL type: delayed, immediate) repeated-measures ANOVA was conducted. Similarly to the global JOLs for the delayed recall test, subjects gave lower global JOLs for list 2 (M = 30.56, SD = 19.23) than list 1 (M = 43.35, SD = 20.41), F (1, 40) = 52.69, $p < .01$, again suggesting that subjects were sensitive to list-level PI. No other effects were significant.

To summarize, control subjects gave higher global JOLs than interference subjects for both the immediate and delayed recall tests. Furthermore, interference subjects gave lower global JOLs for list 2 than list 1 for both immediate and delayed recall tests (subjects never took a delayed recall test for list 1). The global JOL results suggest that interference subjects were sensitive to overall difficulty in encoding list 2.

Appendix D

E2: List 1 and 2 Versions

List 1	Version 1		Version 2		
1	bone	dancer	1	traffic	wood
2	animal	senator	2	dirt	movie
3	throat	wine	3	composer	atom
4	milk	knife	4	dress	wagon
5	wheel	horn	5	cloth	beach
6	column	library	6	baker	queen
7	flower	clerk	7	salt	forest
8	road	prince	8	baby	mayor
9	tool	coast	9	tool	coast
10	prison	tissue	10	prison	tissue
11	wire	fort	11	wire	fort
12	room	bird	12	room	bird
13	crowd	rice	13	crowd	rice
14	tongue	picture	14	tongue	picture
15	handle	corn	15	handle	corn
16	camera	brain	16	camera	brain

List 2	Version 1		Version 2		
1	bone	wine	1	traffic	beach
2	animal	library	2	dirt	queen
3	throat	prince	3	composer	wood
4	milk	clerk	4	dress	movie
5	wheel	senator	5	cloth	atom
6	column	knife	6	baker	wagon
7	flower	horn	7	salt	mayor
8	road	dancer	8	baby	forest
9	traffic	beach	9	bone	wine
10	dirt	queen	10	animal	library
11	composer	wood	11	throat	prince
12	dress	movie	12	milk	clerk
13	cloth	atom	13	wheel	senator
14	baker	wagon	14	column	knife
15	salt	mayor	15	flower	horn
16	baby	forest	16	road	dancer

Appendix E

Memory Beliefs Questionnaire

Question 1:
Participants in an experiment were presented with a list of 60 words with each word presented for 4 seconds each. Their task was to study these words in order to remember them on a recall test later on.

I would like you to estimate how many words the participants were able to recall on average. Your estimate can range from 0 to 60 words. Write down your estimate at the appropriate space at the bottom of the next page.

Please note: Three groups of participants took part in the experiment described. They all studied the same list under the same conditions. For Group A, however, the memory test took place ten minutes later, for Group B it took place two hours later, and for Group C it took place 48 hours later. We have filled in the actual results of Group A. Please use this number to help determine your estimate for the remaining two groups.

"How many words were recalled on average by each group (write a number between 0 and 60 in each space):

After 10 minutes_____**40**_____?

After 2 hours__________?

After 48 hours __________?

Question 2:
A vocabulary learning experiment was conducted with 100 participants. English-speaking participants saw a list of 20 Swahili words paired with their English meanings (e.g, ladha—prayer, zeituni—olives, etc.) one after the other at a rate of 4 seconds per word pair. Their task was to study these pairs so that when presented later with the first, Swahili word ("ladha" in the example), they would be able to recall the second, English word that went with it ("prayer" in the example). In this way, the first word served as a "hint" and the second word served as the "response."

Fifty of the 100 participants learned **two** lists of 20 word pairs (the 2-list group), with the second list right after the other. For this group, half (10) of the word pairs in the second list paired some of the old Swahili words with their second English meanings. For example, if "ladha--prayer" appeared in the first list, then "ladha--flavor" would appear in the second list. The other 10 word pairs on the second list were completely new Swahili-English word pairs (e.g., sanda--dipper).

To help illustrate the details further, the 2 lists with 20 word pairs are listed on the next page. You are not required to study the word pairs, but only to estimate how many word pairs the participants recalled. Keep in mind that the participants had to recall the English word when presented with the Swahili word.
The other 50 participants, in the 1-list-only group, learned the same second list but did **not** learn the first list.

Please note: For the 2-list group, you are estimating their recall for the second list ONLY. Again, the list that the 1-list-only group learned **is the same list** as the second list that the 2-list group learned. Below we have filled in the 1-list-only group's actual performance. Please make an estimate for the other group.

A) On average, what percent of the 20 word pairs will the 1-list group have correctly learned:

1-list-only group ___**60%**____?

B) On average, what percent of the List 2 word pairs out of 10, will be learned correctly by the 2-list group, for word pairs where the Swahili word from List 1 was paired with a new, second meaning on List 2?

Please write down an estimate percentage: _________%.

C) On average, what percent of the List 2 word pairs out of 10 will be learned correctly by the 2-list group, for word pairs that presented completely new Swahili words with their English translation?

Please write down an estimate percentage: __________%

Appendix F

Figures

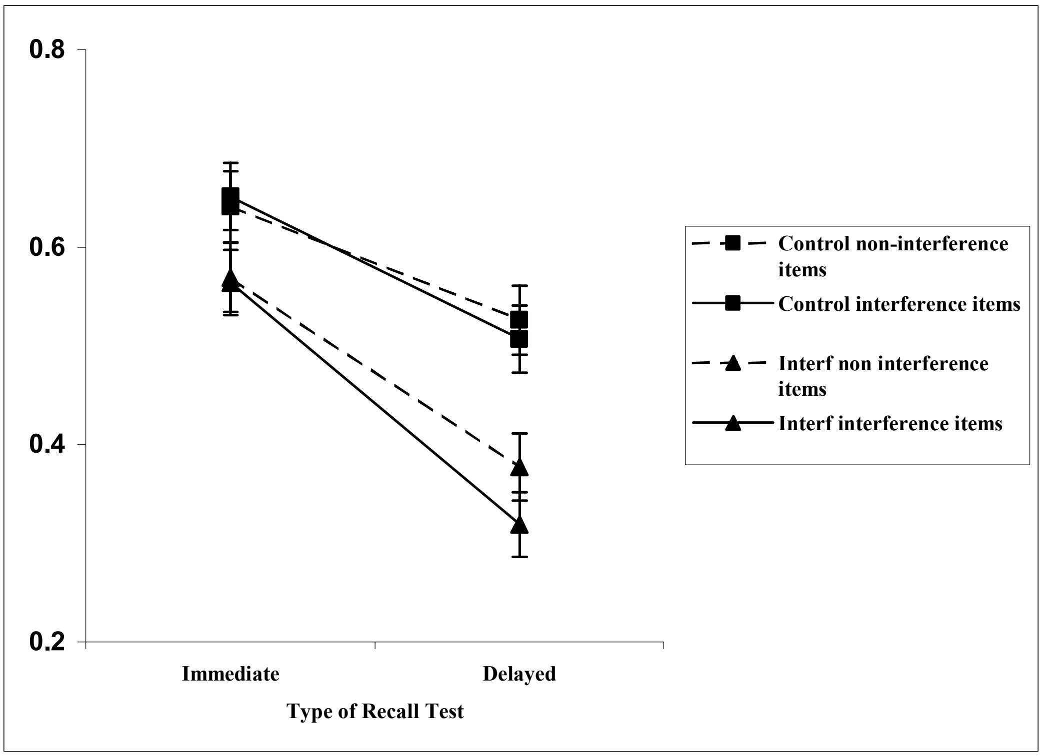

Figure 1. E1: Comparing interference and control condition subjects on mean proportion correct on immediate and delayed recall tests.

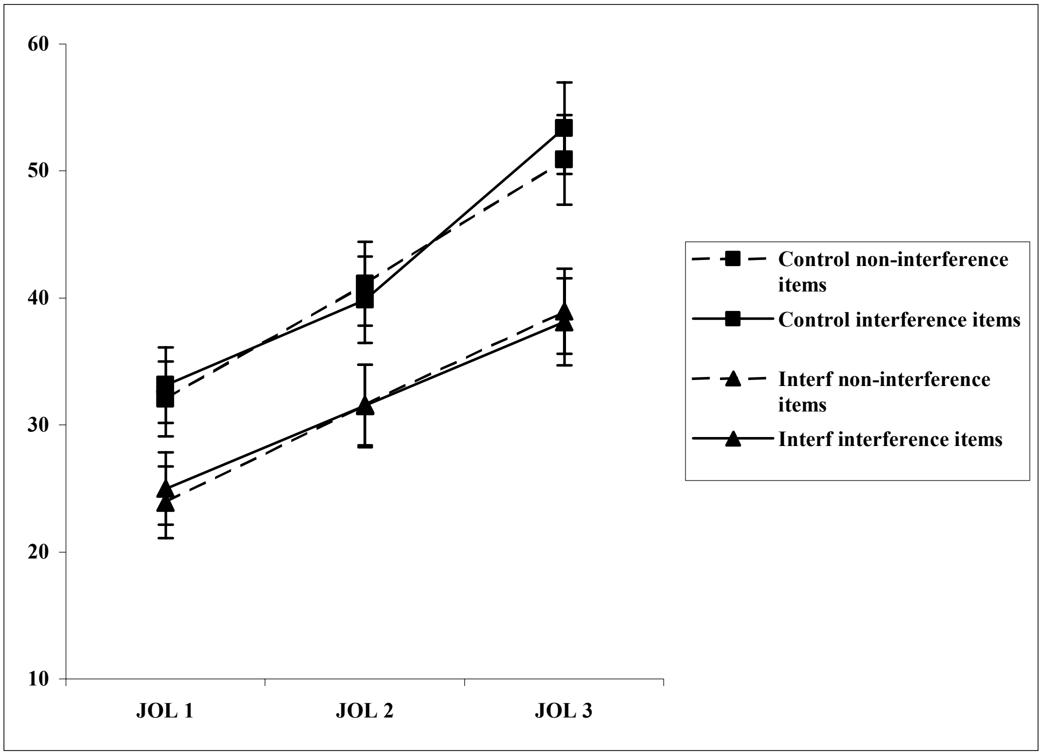

Figure 2. E1: Comparing the interference and control conditions on the overall magnitude of specific-item JOLs for list 2 across learning trials.

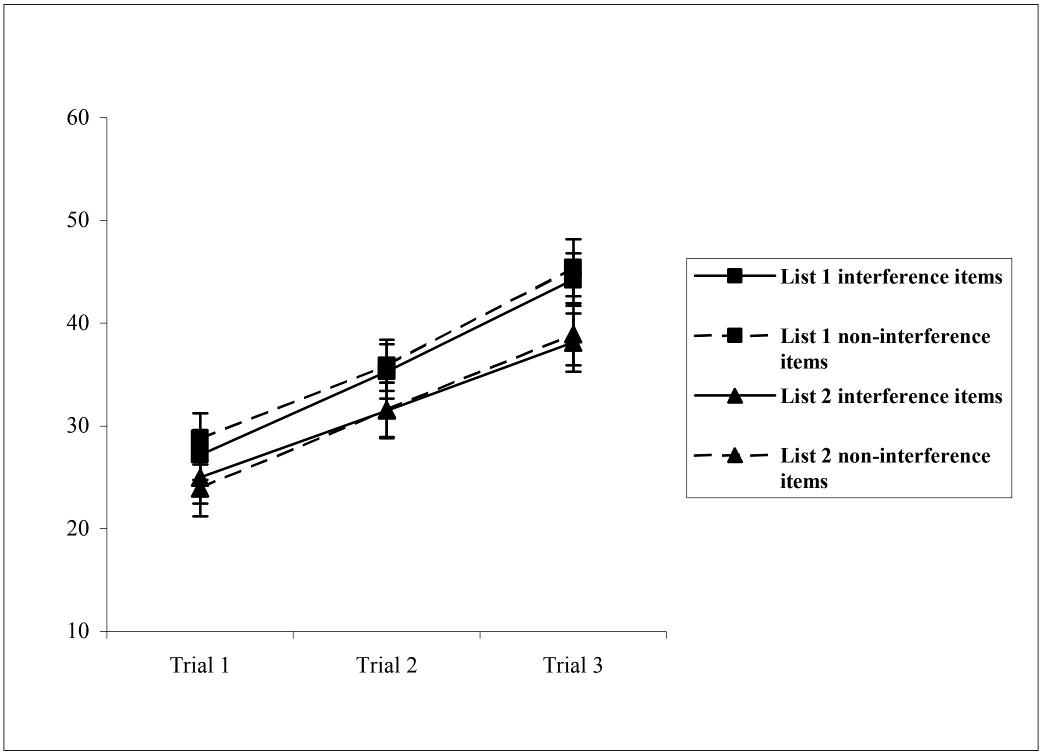

Figure 3. E1: Comparing interference subjects' List 1 and List 2 JOLs across learning trials.

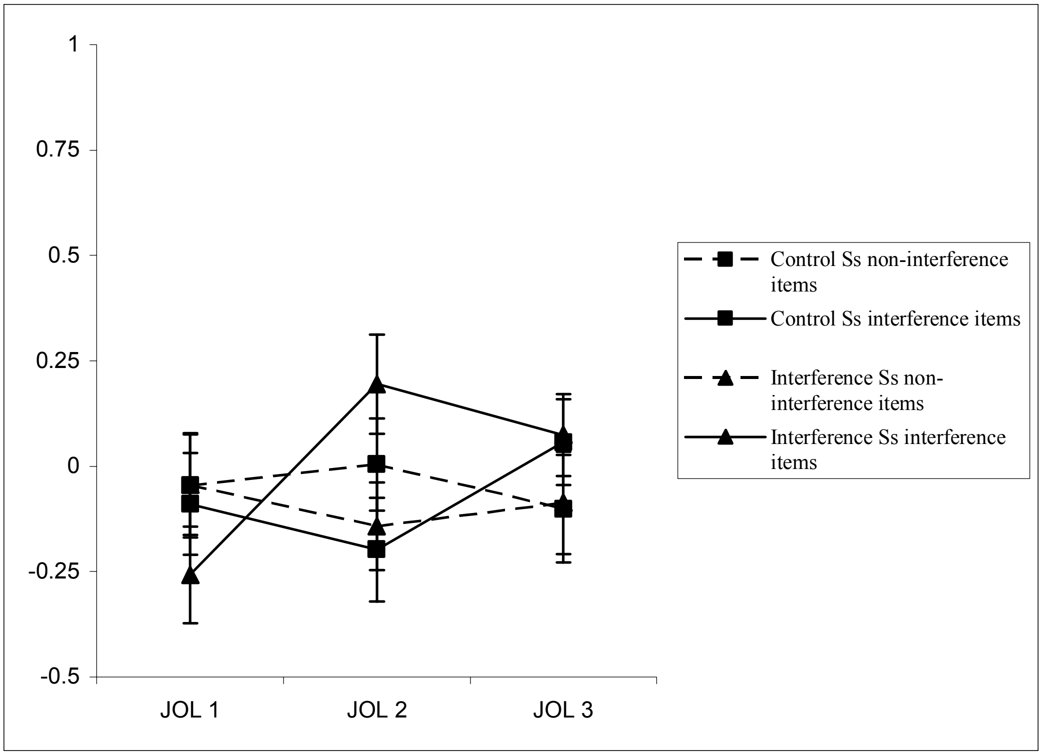

Figure 4. E1: Mean gamma correlations between list 2 JOLs and delayed recall across learning trials

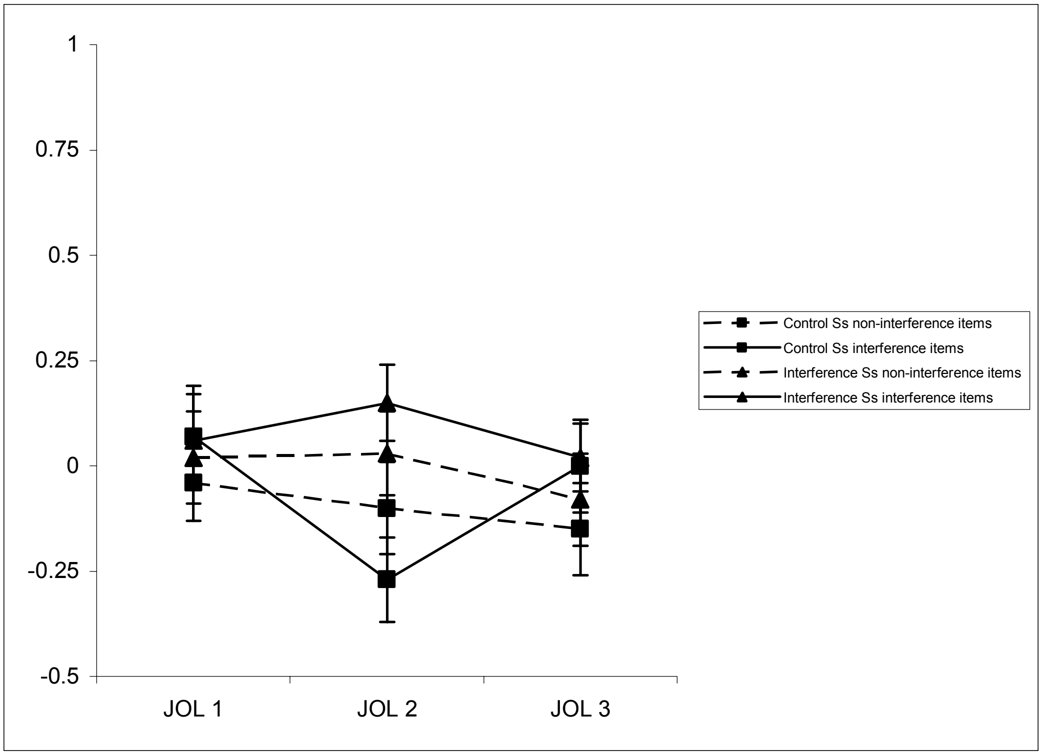

Figure 5. E1: Mean gamma correlations between list 2 JOLs and immediate recall across learning trials

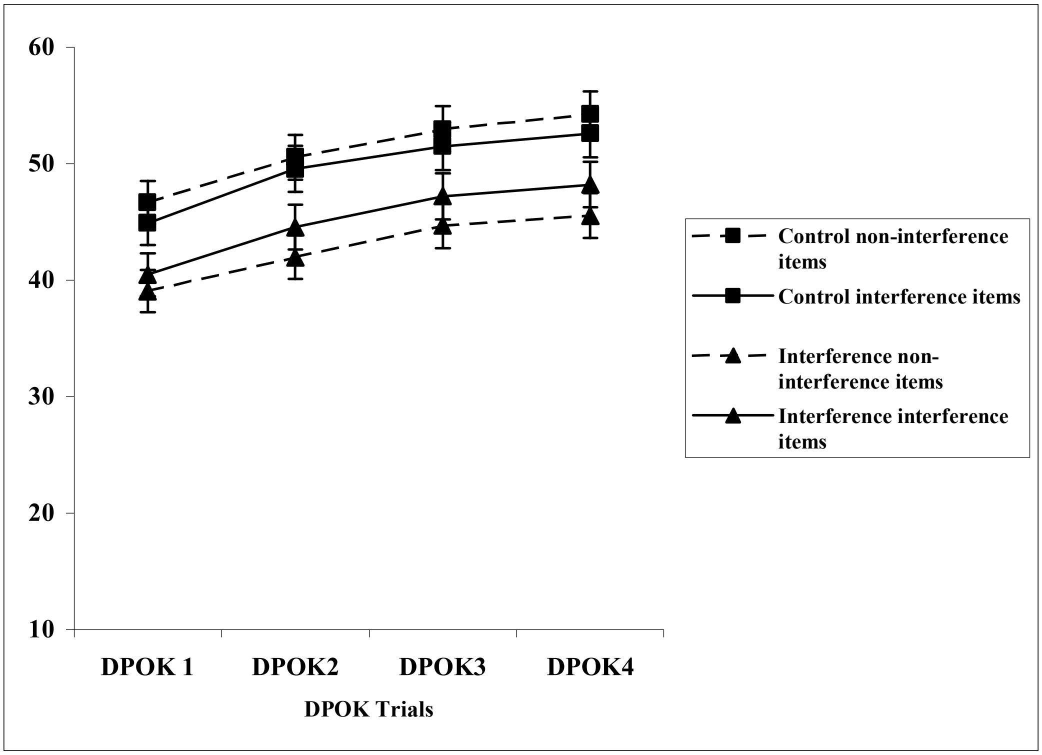

Figure 6. E1: Comparing interference and control subjects on mean DPOK judgment magnitude for list 2 word pairs.

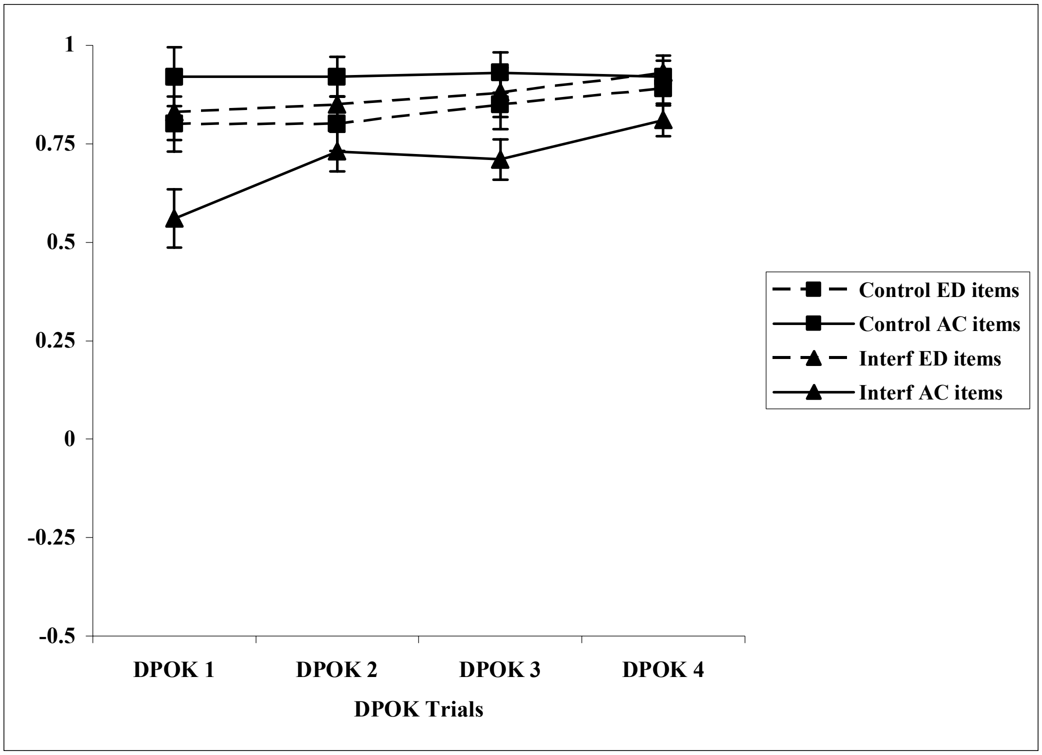

Figure 7. E1: Comparing interference and control subjects' gamma correlations between DPOK judgments and delayed recall for interference and non-interference items.

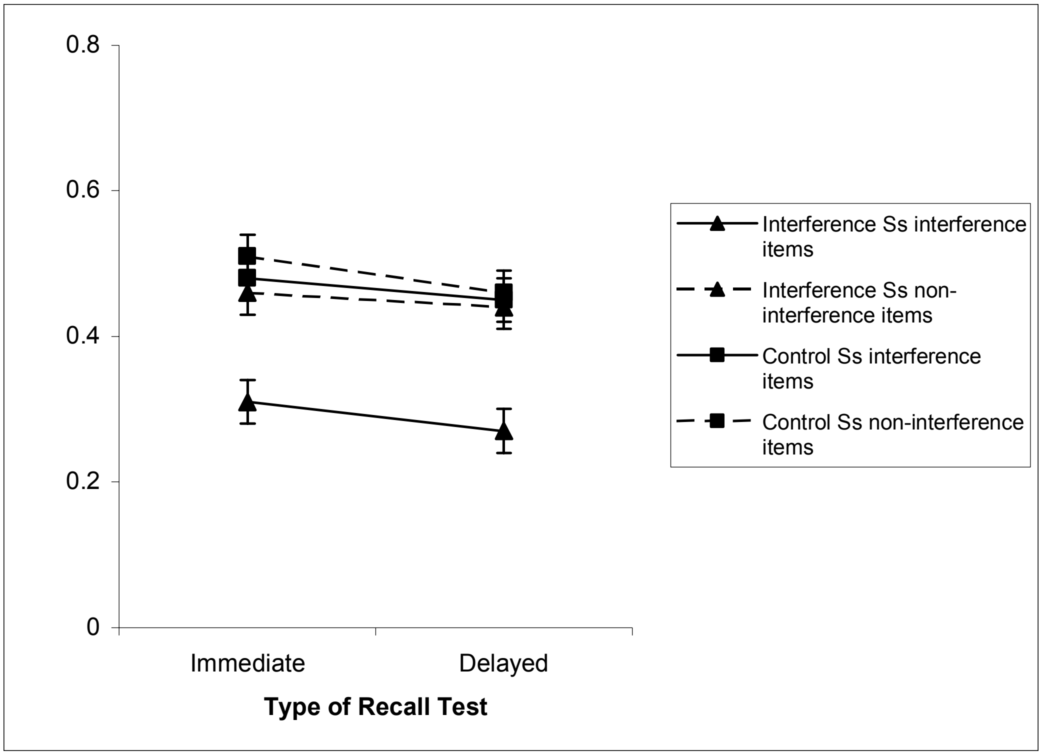

Figure 8. E2: Comparing interference and control condition subjects on mean proportion correct on immediate and delayed recall tests.

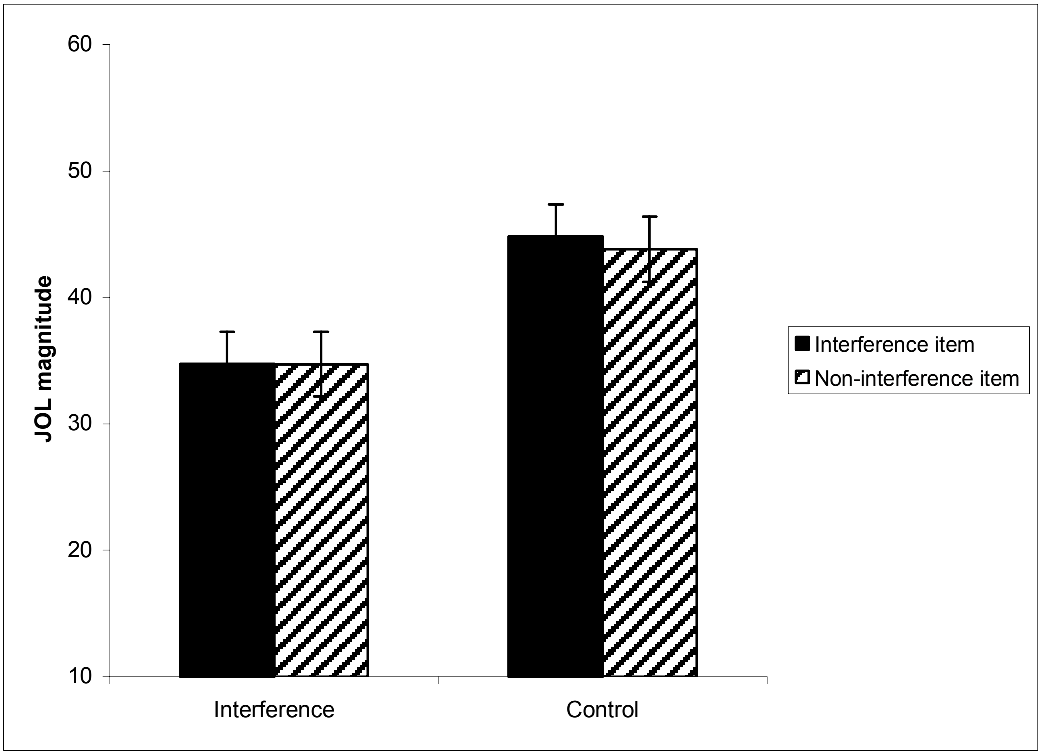

Figure 9. E2: Comparing the interference and control conditions on the overall magnitude of specific-item JOLs for list 2.

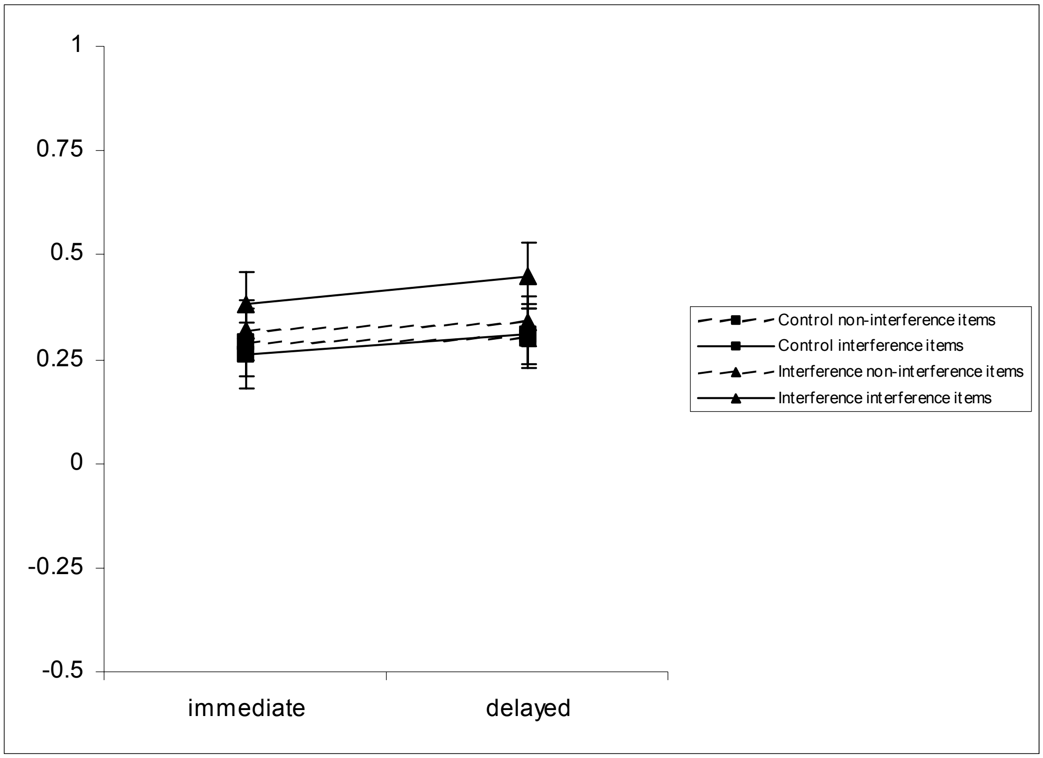

Figure 10. E2: Mean gamma correlations between list 2 JOLs and immediate and delayed recall for interference and control subjects

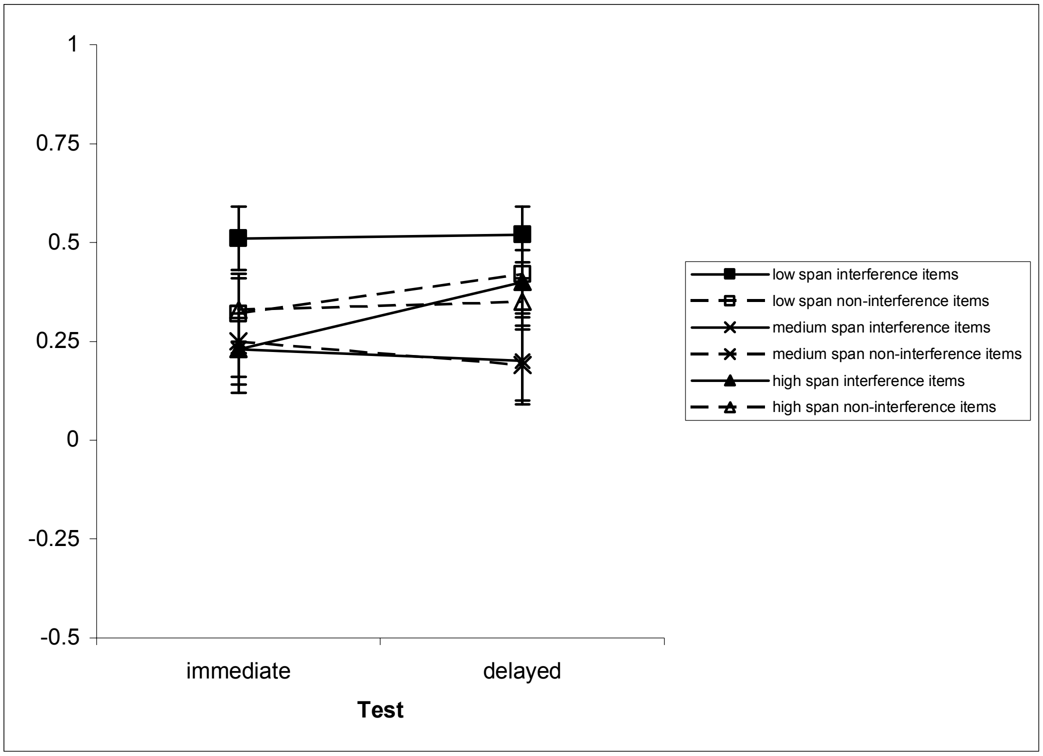

Figure 11. E2: Mean gamma correlations between list 2 JOLs and immediate and delayed recall for high, medium, and low WM interference and control subjects

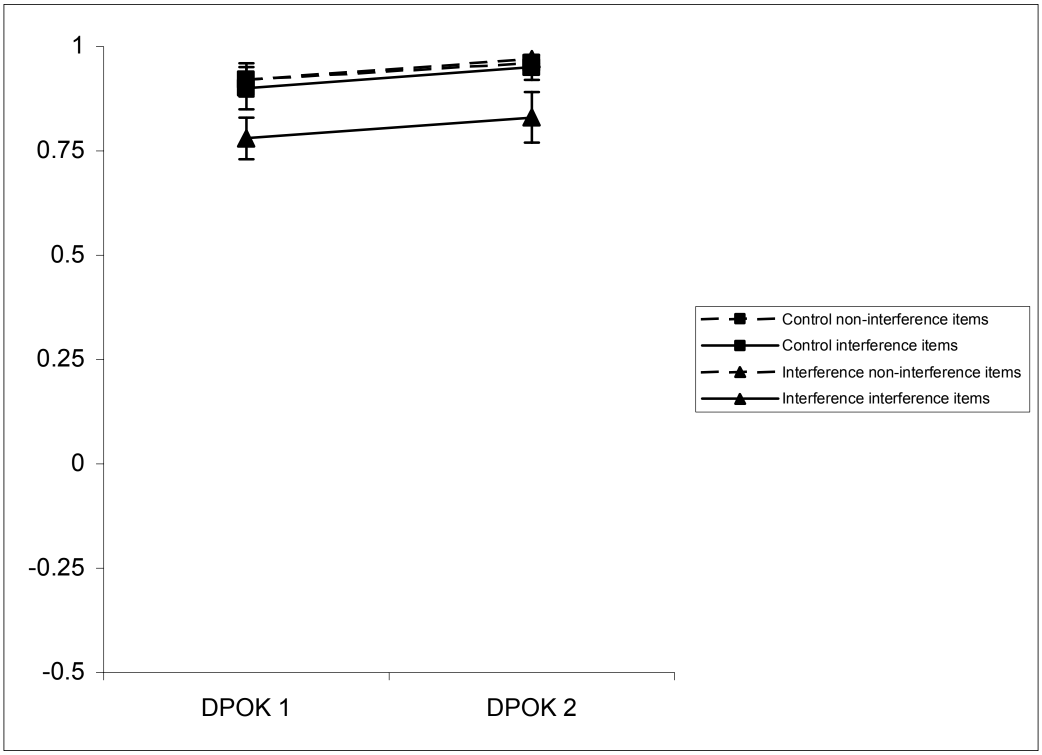

Figure 12. E2: Mean gamma correlations between DPOKs and delayed recall for interference and control subjects

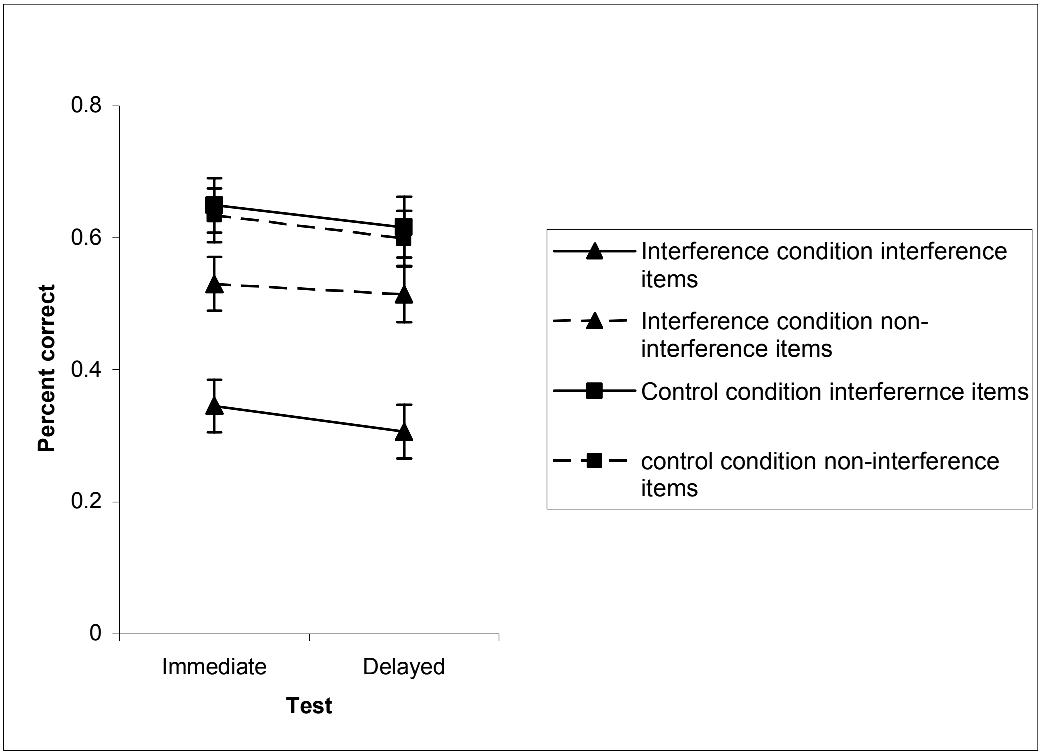

Figure 13. E3: Comparing interference and control condition subjects on mean proportion correct on immediate and delayed recall tests.

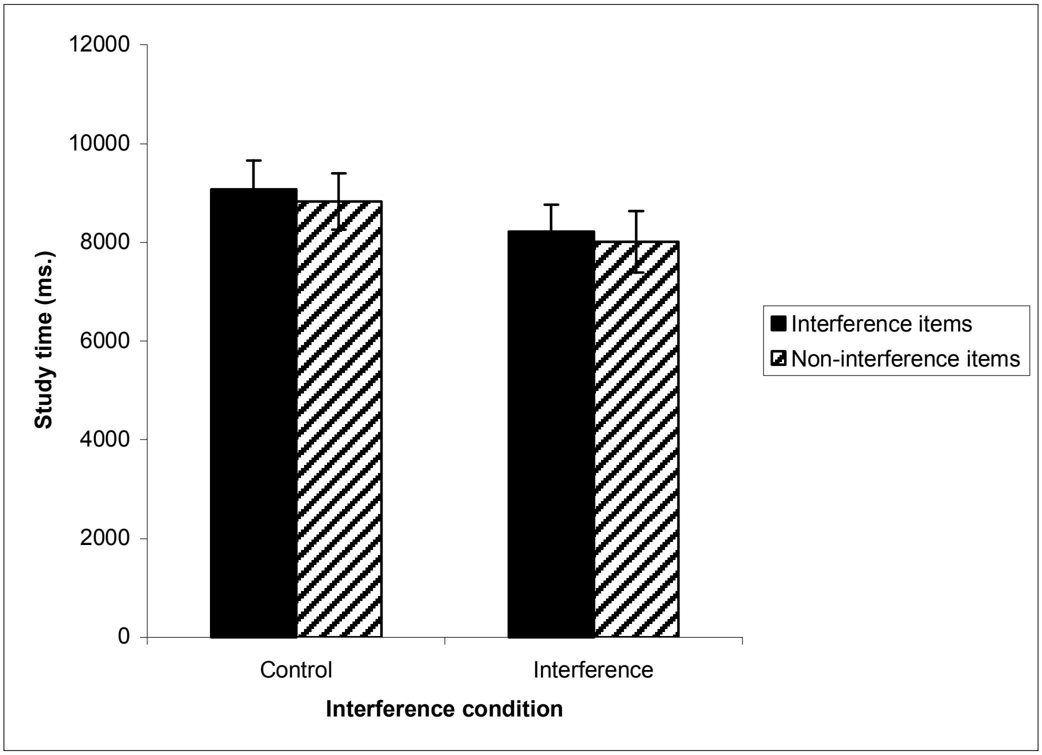

Figure 14. E3: Comparing interference and control subjects' list 2 study time for interference and non-interference word pairs.

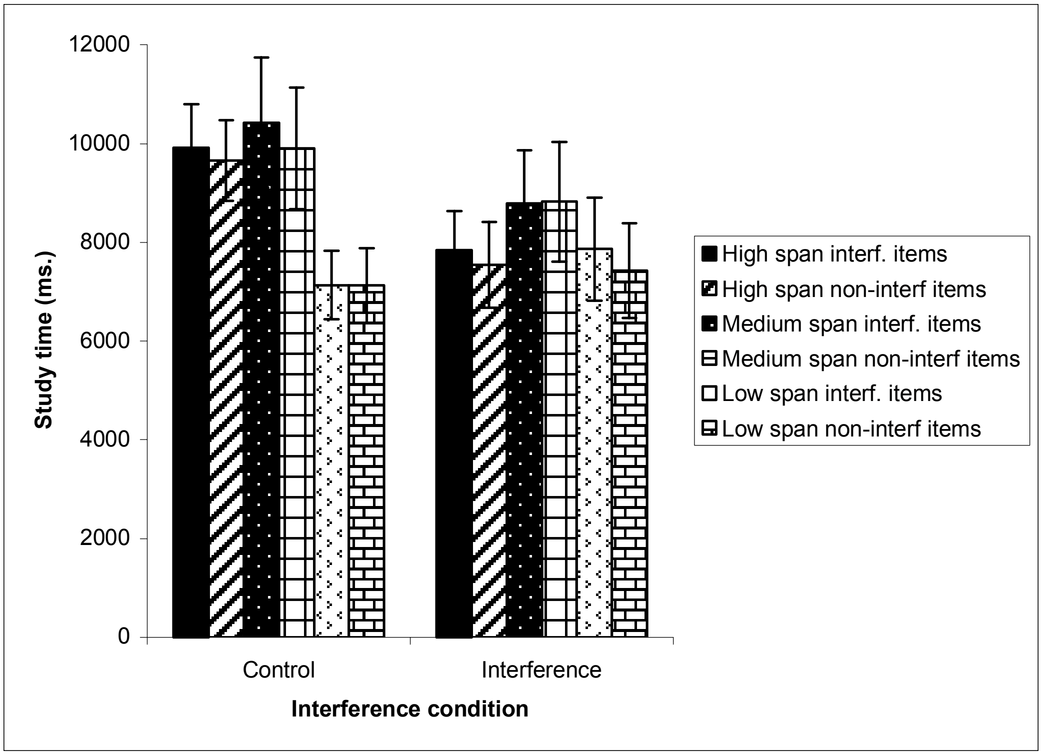

Figure 15. E3: Comparing high, medium, and low span interference and control subjects' list 2 study time for interference and non-interference word pairs.

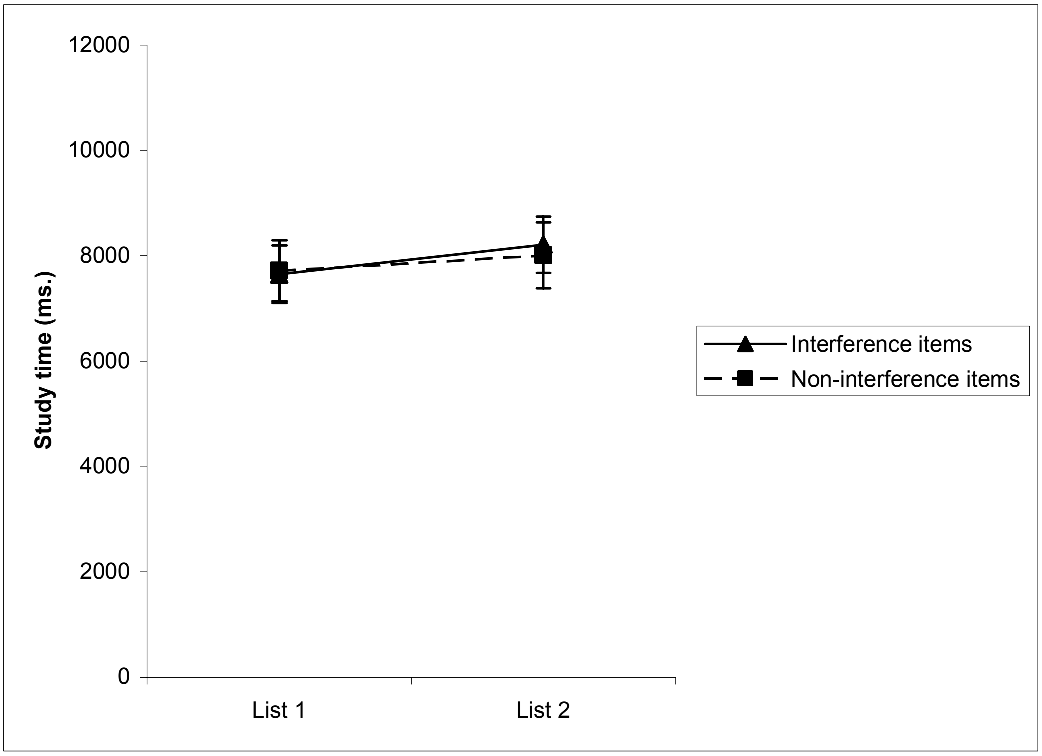

Figure 16. E3: Comparing interference subjects' list 1 and 2 study time for interference and non-interference word pairs.

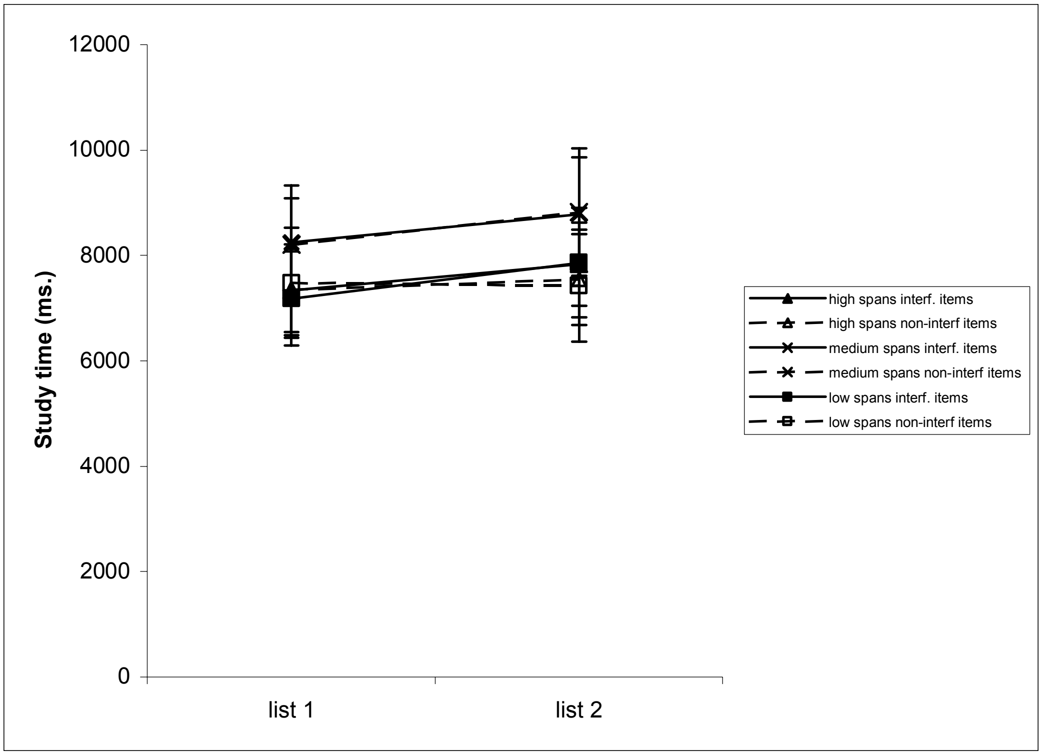

Figure 17. E3: Comparing high, medium, and low span interference subjects, list 1 and 2 study time for interference and non-interference word pairs.

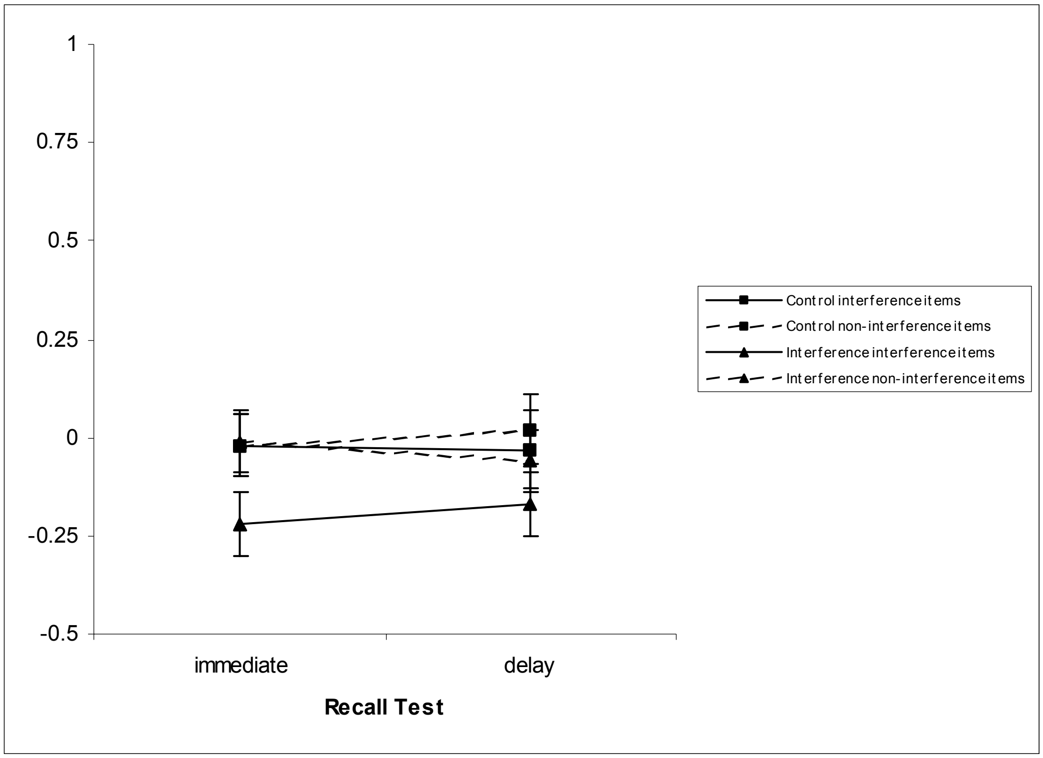

Figure 18. E3: Comparing control and interference subjects' gamma correlations between study time and immediate and delayed recall.

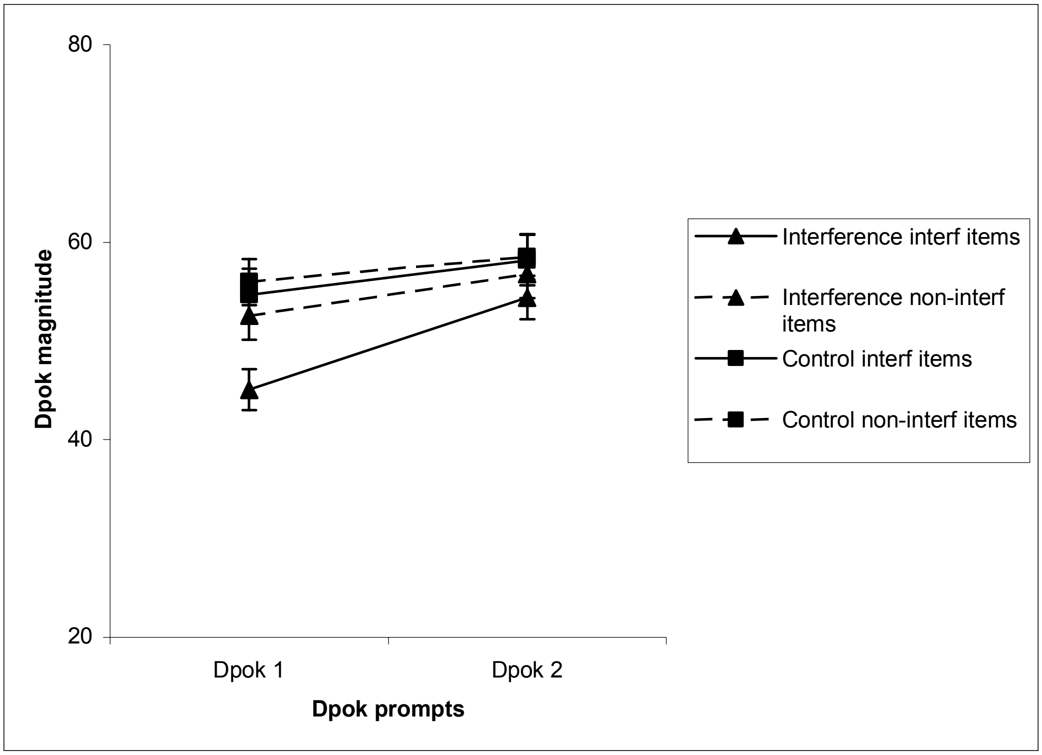

Figure 19. E3: Comparing control and interference subjects' DPOK judgments.

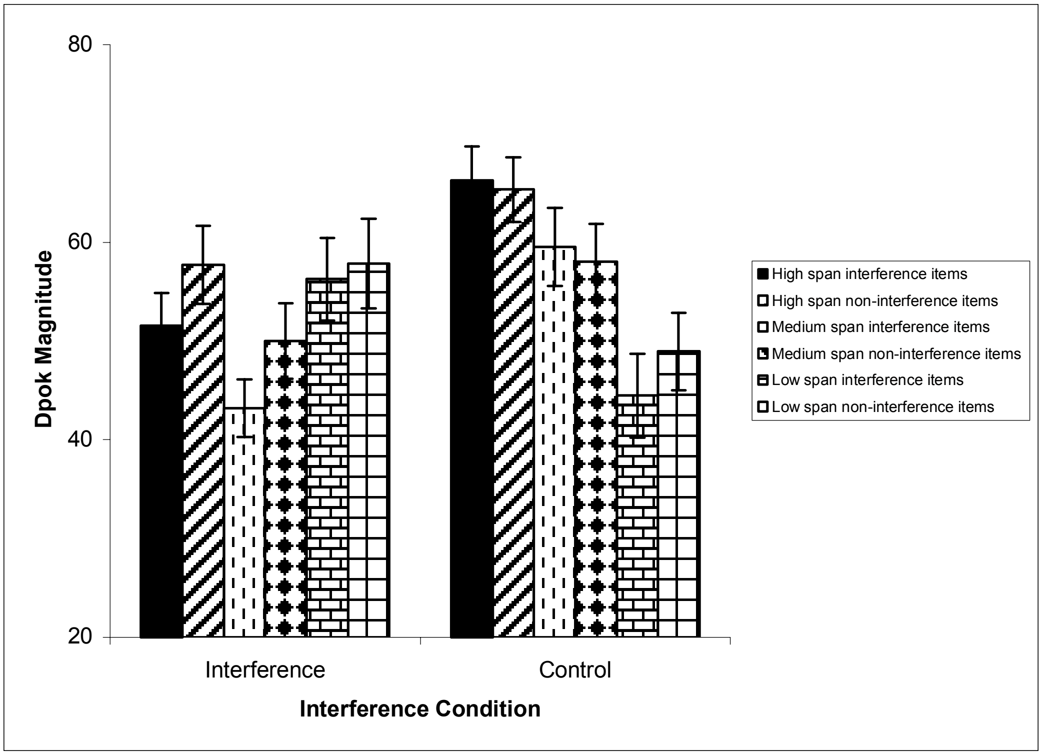

Figure 20. E3: Comparing high, medium, and low span subjects' DPOK magnitude for interference and non-interference items.

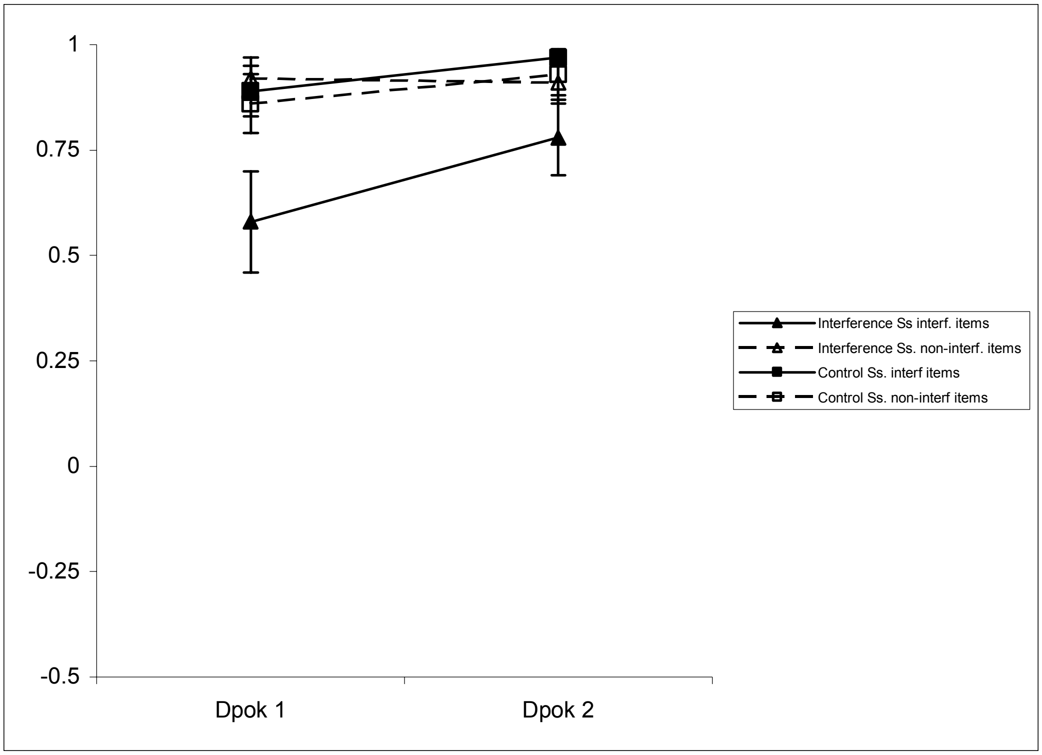

Figure 21. E3: Comparing control and interference subjects' gamma correlations between DPOKs and delayed recall for interference and non-interference items.